DOCUMENTOS DE TRABAJO

Evaluation of Short Run Inflation Forecasts in Chile

Pablo Pincheira
Roberto Álvarez

N.º 674 Agosto 2012

BANCO CENTRAL DE CHILE

DOCUMENTOS DE TRABAJO

Evaluation of Short Run Inflation Forecasts in Chile

Pablo Pincheira
Roberto Álvarez

N.º 674 Agosto 2012

BANCO CENTRAL DE CHILE

BANCO CENTRAL DE CHILE

CENTRAL BANK OF CHILE

Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180, Santiago, Chile
Teléfono: (56-2) 3882475; Fax: (56-2) 3882231

Documento de Trabajo
N° 674

Working Paper
N° 674

EVALUATION OF SHORT RUN INFLATION FORECASTS IN CHILE*

Pablo Pincheira
Banco Central de Chile

Roberto Álvarez
Banco Central de Chile

Abstract
The Central Bank of Chile builds inflation forecasts for several time horizons and using various methodologies. In this paper, we analyze one of these series of short-term inflation forecasts, which we call Auxiliary Inflation Forecasts (AIF), comparing them to forecasts made by private analysts and to forecasts built from simple time-series models. We also evaluate the AIF using encompassing tests and bias and weak efficiency tests. Our aim is to answer two linked questions: first, which is the best forecast series under a specific loss function? and second, are the differences of accuracy between any two series of forecasts totally explained by the differences in the information sets from which they have been built? Our results indicate that the AIF behave extremely well at one- and two-month horizons, but they are less adequate at longer horizons.

Resumen
El Banco Central de Chile construye pronósticos de inflación de acuerdo con múltiples metodologías y horizontes. En este artículo analizamos uno de estos pronósticos de inflación de corto plazo, que son denominados Pronósticos Auxiliares de Inflación (PAI), comparándolos con proyecciones realizadas por analistas privados y por simples modelos de series de tiempo. También evaluamos los PAI utilizando tests de encompasamiento, sesgo y eficiencia débil. Nuestro objetivo es responder dos preguntas interconectadas: primero, ¿cuál de las series es la más precisa bajo una determinada función de pérdida? y, segundo, ¿pueden ser totalmente explicadas las diferencias entre dos series de pronósticos a partir de las diferencias existentes entre los conjuntos de información sobre los que se construyen? Nuestros resultados indican que los PAI se comportan extremadamente bien en horizontes de uno y dos meses, pero tienen un desempeño más moderado en horizontes largos.

---

* We are grateful to Álvaro García, María Ignacia Valencia and Mariana García for efficient research assistance. We also thank Klaus Schmidt-Hebbel, Macarena García, Andrea Sánchez and Felipe Córdova for their generous conceptual contributions and Michael Pedersen for his valuable comments and for allowing us the use of the auxiliary inflation forecasts. Our work has also benefited from useful discussions with Pablo García and the opinions expressed at the Central Bank's workshop on monetary policy. All remaining errors are ours. Correspondence: Agustinas 1180, Santiago, Chile, Phone: 56-2-670-2874, Fax: 56-2-670-2836. E-mail: ppinchei@bcentral.cl; ralvarez@bcentral.cl;

1. Introduction

Central banks conduct monetary policy based on multiple pieces of information that enable them to envisage scenarios about the possible evolution of the economy. Some of these pieces of information are in the form of sets of inflation forecasts at several horizons. It is reasonable to think that a successful monetary policy depends at least in part on the quality of these inflation forecasts.

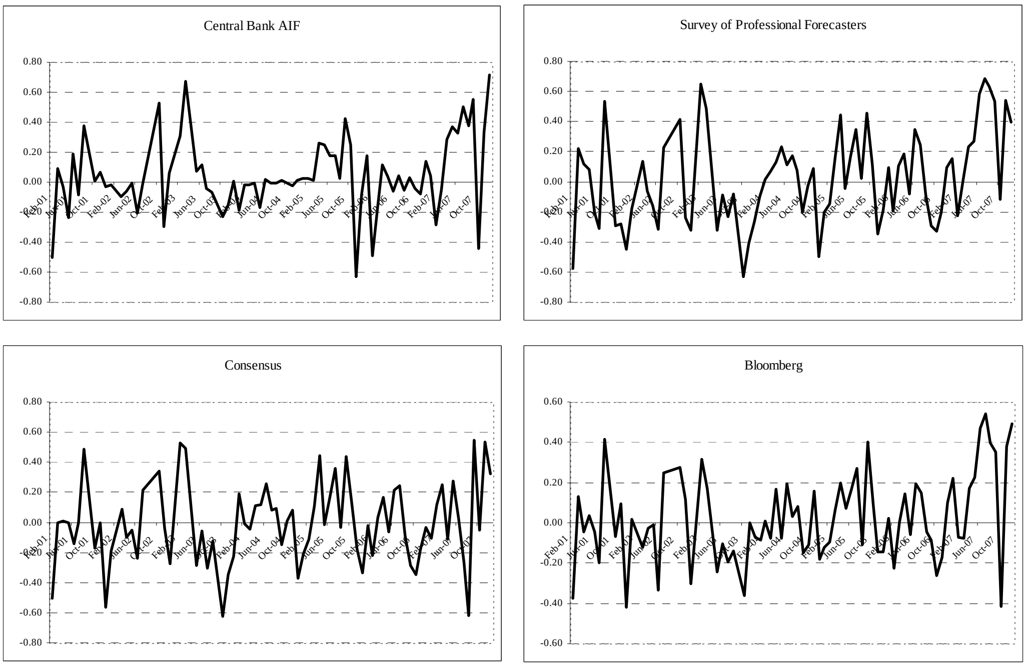
On the basis of this reflection, in this paper we evaluate one of the series of short-term inflation forecasts used by the Central Bank of Chile, which we will call Auxiliary Inflation Forecasts (AIF). These forecasts correspond to six series of predictions over horizons going from one to six months ahead.

We evaluate these forecasts with the aim of answering two linked questions. First, which is the best set of forecasts under a specific loss function? Second, are the differences of accuracy between any two series of forecasts totally explained by the differences in the set of information from which they have been built? In other words, if the series of forecasts built by analyst A is more accurate than the series built by analyst B, and the series of A was built on a broader set of information than that used by B, can we claim that the difference in accuracy is totally due to differences in information? This last question represents the main conceptual contribution of this paper, and it is particularly relevant because the available inflation forecasts are usually built at different moments in time. Therefore we expect a forecast built on more information to be more accurate than a forecast built on less information. We propose a simple scheme of analysis based on a theoretical context to determine whether differences in quality are due to initial differences in the available information at the moment of prediction.

To motivate our first question, Figures 1 and 2 show the evolution of forecast errors, defined as the difference between effective inflation and predicted inflation. Figure 1 shows the prediction error of the AIFs and other analysts over a one-month horizon. As expected, there are substantial differences between the forecasts and actual inflation[1].

This simple evidence reveals that the magnitude of one-month ahead prediction errors has changed over time and can reach economically significant values with occasional peaks of 70 basis points[2].

Figure 1: **1-Month Ahead Forecast Errors, 2001-2007**



[1] The evidence is similar no matter if the median or mean forecast is taken when information comes from several analysts, as in the case of the Survey of Professional Forecasters and Bloomberg.

[2] This means that if the predicted inflation, expressed as its change in 12 months, were 3.0%, effective inflation would reach about 3.7%.

Figure 2: 6-Month Ahead Forecast Errors, 2001-2007

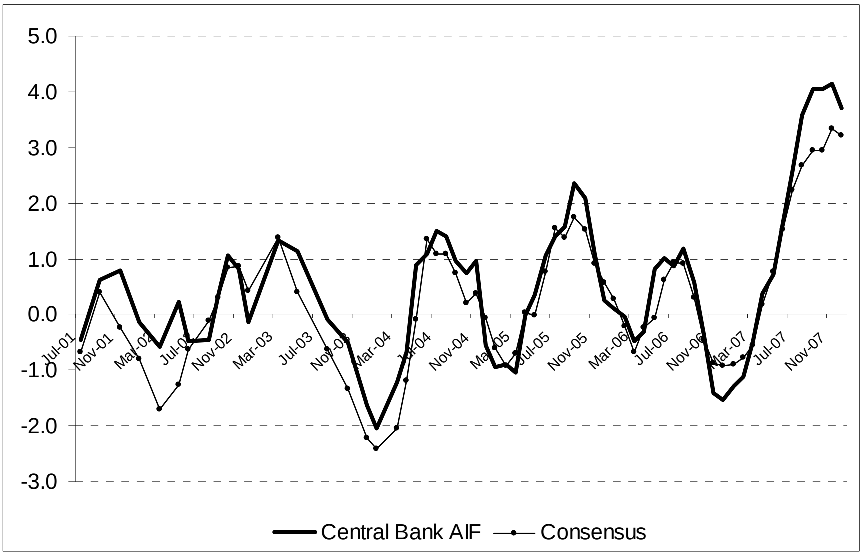


Figure 2 shows six-months ahead prediction errors for inflation made by the AIFs and Consensus Forecast. Three aspects are worthy of mention. First, forecasting errors at the end of the sample period are the largest for both the AIFs and Consensus Forecast. Secondly, both sources' errors present relatively similar behavior and display a downward bias. Thirdly, AIF errors have been somewhat larger than those of Consensus Forecast in the last few months of high inflation.

Based on this evidence, several empirical questions arise: are the AIFs better or worse than forecasts made by other analysts? Is there any bias in these forecasting errors? Are the AIFs efficient in incorporating all available data? Would the AIFs be improved by somehow incorporating the forecasts of private analysts?

We answer these questions in three stages. First, we compare the AIFs with alternative forecasts from private analysts and simple time series models. This gives us a certain objectivity in assessing the forecasts because using only other analysts' forecasts as benchmarks involves the risk of using lax criteria since we cannot be sure that their predictions are of a high quality. Secondly, we evaluate the quality of the AIFs according to bias and weak efficiency criteria. Thirdly, we make an encompassing analysis to detect whether AIFs could benefit from other analysts' forecasts.

Finally, we would like to point out an interesting additional finding from this research. Unlike the evidence of a vast amount of forecasting combination studies, the simple average of a set of forecasts from the Central Bank of Chile's Survey of Professional Forecasters (SPF) behaves quite discreetly compared to the individual forecasts. This result is valuable in that it provides evidence in conflict with to the so-called "combination puzzle" (see Stock and Watson, 2004; Smith and Wallis, 2008).

This paper is structured as follows: the next section presents a review of the literature; section 3 describes the data; section 4 presents the methodology; section 5 assesses the relative performance of AIF according to various criteria, comparing them with private analysts' forecasts and those derived from simple time series models; section 6 gives bias and efficiency results; section 7 shows the encompassing analysis; and finally we give our conclusions and a brief summary of the article in section 8.

2. Literature Review

Several of the empirical questions we raised in the introduction have already been tackled to some extent in the literature on predictive ability. Most of these works use similar methodology to ours. In general, they consider some measure of prediction error (mean squared prediction error, absolute prediction error, etc.) and they evaluate whether any significant difference exists between the diverse series of forecasts. Almost all the works also analyse the bias and efficiency of the forecasts and compare their performance with some simple econometric or time series models.

Andersson et al (2007) carry out a similar exercise to ours in their assessment of the relative performance of the Riksbank. In general, they find that the Swedish Bank's forecasts are more accurate than the benchmark they define (forecasts provided by the National Institute of

Economic Research) but the differences are not statistically significant. Moreover, their results suggest that the Swedish Central Bank performs quite well compared to Consensus' forecasts.

In a different context, Capistrán and López-Moctezuma (2008) evaluate inflation, exchange rate, GDP growth and interest rate forecasts produced by the Bank of Mexico's survey of professional forecasters, based on notions of efficiency and accuracy. For these four series and over most of the horizons, they find evidence about macroeconomic information, available at the moment of prediction, that is not really used by these forecasters. In terms of predictive accuracy, the evidence shows that the survey's ability to predict inflation is more accurate than an autoregressive model for almost all the horizons considered and that the exchange rate forecasts one and two-months ahead are better than those provided by a random walk model.

Another interesting work for several economies by Oller and Barot (2000) compares growth and inflation forecasts from the OECD and research institutes for 13 countries. Their findings show that there are not significant differences between the prediction errors of these two sources of forecasts.

In another article, Loungani (2001) evaluates the prediction errors of Consensus Forecast with regard to GDP growth in several developed and developing countries. She finds some evidence of inefficiency and overestimation and shows a high correlation between the forecasts of international institutions (World Bank, IMF and OECD).

More recently, Bowles et al (2007) evaluate the European Central Bank Survey of Professional Forecasters' predictions over eight years. One of their conclusions is that the respondents systematically underestimate inflation throughout the evaluation period[3].

---

[3] On similar lines, Ang, Bekaert and Wei (2007) in an out-of-sample exercise evaluate four traditional methods of predicting inflation in the United States and conclude that survey-based measures achieve the best results. Croushore's (1998) analysis of various US inflation forecasting sources also interestingly finds that the errors of prediction have tended to diminish over time.

Some research has also been carried out in Chile. In particular, the recent work of Bentancor and Pincheira (2008) shows that the inflation forecasts from the Central Bank of Chile's Survey of Professional Forecasters (SPF) display a significant downward bias and excess of autocorrelation in the second half of the sample period. By correcting the bias and autocorrelation in an out-of-sample exercise the authors achieve significant reduction in mean squared prediction error (MSPE) and bias.

Also for Chile, Albagli et al (2003) assess growth and inflation prediction errors of the Central Bank of Chile, private analysts and other central banks. With regard to inflation, they show that from 2000 to 2002, the Central Bank of Chile's errors were considerably smaller than those of private analysts[4].

Finally, Chumacero (2001) analyses private forecasters' estimates of GDP growth rates during the period 1986-1998 for Chile. His results show that forecasters systematically underestimate the true growth rate of the economy.

Thus, this brief and selective review of the evidence shows that it is not uncommon to find systematic errors in both public and private forecasters' predictions and there are plenty of cases of forecasting inefficiency. This is true with regard to forecasts of various macroeconomic variables and for developed and developing economies. Interestingly enough, none of the articles we mention ask whether the differences in accuracy found between two forecasters are due to the potential differences in the sets of information they use to make their forecasts. In the following sections we will see how our results fit in with the existing literature.

[4] The root mean squared error of the Bank's forecasts was 28% less than that of Consensus and 41% less than that of the SPF.

3. Source of data

We use monthly data for the period from January 2001 to December 2007. We consider one month to six months ahead forecasts. This period is chosen due to series availability: There were not AIFs series prior to January 2001.

We consider 4 sources of alternative forecasts: The first is the SPF carried out periodically by the Central Bank of Chile. Each individual analyst's inflation forecast is gathered at the beginning of the month for which the forecasts are made. This survey includes forecasts with horizons of one and three months ahead. The second data source is Consensus Forecast which makes a mid-month survey which implies that forecasts for the current month are a half-step ahead forecast. This data source provides forecasts at horizons from one to six months ahead but only gives the respondents' forecasts in the aggregate.

A third data base is built from Bloomberg's periodic surveys. Most of these forecasts are published between one day and one week prior to the publication of the actual inflation rate. Although this horizon is much shorter than one month, we treat them as if they were one-month ahead forecasts. In this case the information is also from each respondent.

The fourth data source is built from the automatic selection of an autoregressive moving average model (ARMA(p,q)) estimated period by period with rolling windows.

The forecasts database has some missing observations. This is partly because a private analyst rarely provides forecasts for every single month during the sample period. We do not deal with missing observations in a formal statistical way but we do attempt to mitigate it by: firstly, ensuring that we compare the AIFs with private forecasts that have at least 50% of the observations that the AIFs have; and, secondly, by restricting our data to only include observations corresponding to dates on which both the AIFs and the private forecasts do not display missing values[5]. For example, to evaluate the mean prediction error, if the AIFs have data for January, February and March 2005, but the private forecasts only have data for January and March 2005, we calculate the AIF mean prediction error omitting the data for February 2005[6]. In this way, we avoid a potential bias associated with the possibility of strategic omission when the inflation scenario looks uncertain.

4 Methodology

As stated in the introduction, the purpose of this work is to assess in multiple dimensions the accuracy and quality of the AIFs. First, we compare the AIFs with alternative forecasts from private analysts and simple time series models. This give us a certain objectivity in assessing the forecasts because using other analysts' forecasts as benchmarks involves the risk of using lax criteria since we cannot be sure that these are high-quality predictions. Secondly, we assess the quality of the AIFs according to bias and weak efficiency criteria. Thirdly, we make an encompassing analysis to detect whether AIFs could benefit from other analysts' forecasts. The following subsections describe these three stages in detail and the way in which we treat the heterogeneity of the available information sets.

4.1 Comparison with private forecasters and simple time series methods

Here we compare AIF performance with private forecasters' using Mean Squared Prediction Error (MSPE) as a measure of predictive accuracy[7]. The MSPE is defined as follows:

$$MSPE(e) = E(e^2)$$

where e denotes the prediction error, defined as the actual minus the predicted value. For this loss function we evaluate predictive ability using 12-month log change in inflation.

[5] This is to avoid making comparisons with only a few observations which therefore might not adequately represent the relative behavior of the two series of forecasts.

[6] The AIFs can therefore present different averages when they are compared to different analysts.

[7] Although most literature uses error measures drawn from statistics, McCulloch and Rossi (1990), Leitch and Tanner (1991) and West, Edison and Cho (1993) use economic-based measures. This is the case of evaluations where the loss functions are associated with economic criteria such as profits or measures of welfare. This kind of evaluation goes beyond our objective.

Unless stated otherwise, the MSPE comparative graphs are reported as ratios with respect to the value obtained for the AIFs. Thus a ratio less than 1 indicates that the AIFs have been outperformed in predictive accuracy. On the contrary, a ratio higher than 1 indicates that the AIFs have performed better than the corresponding analyst or group of analyists.

To determine whether the potential MSPE differences are systematic or random we need to use statistical inference. We follow the evaluation framework proposed by Giacomini and White (2006). Although in practice and under specific operational assumptions this paradigm can be reduced to one very similar to that proposed by Diebold and Mariano (1995) and West (1996), there are relevant conceptual differences. In fact, the tests proposed by Giacomini and White (2006) aim to evaluate a forecasting method and not a forecasting model. This distinction, albeit subtle, is very relevant to our work because the observations obtained are inflation forecasts that are not necessarily associated with specific models.

The version of Giacomini and White (2006) tests that we use in this work use a statistic originally attributed to Diebold and Mariano (1995) and West (1996) with one consideration: no correction is made for parameter uncertainty since we do not want to evaluate a model with population parameters but a forecasting method.

The following statistic is built:

$$t_{n(h)} = \frac{\Delta\overline{L_{n(h)}}}{\sigma_{n(h)} / \sqrt{n(h)}}$$

$$with$$

$$\Delta\overline{L_{n(h)}} = \frac{1}{n(h)} \sum_{t=1}^{n(h)} \Delta L_t$$

in which *h* represents the forecast horizon, *n(h)* represents the number of forecasts for the corresponding horizon, *ΔL* represents the loss differential between the AIFs and one specific analyst's forecast and $\sigma_{n(h)}$ is a HAC estimator of the asymptotic standard deviation of the statistic numerator $t_{n(h)}$ multiplied by root square of n(h). For all practical effects, we proceed using a HAC estimate according to Newey and West (1987) with automatic lag selection according to Newey and West (1994).

Under the assumptions described in Giacomini and White (2006), the $t_{n(h)}$ statistic is asymptotically normal under the null hypothesis of equal predictive ability.

There is one important limitation when comparing the AIFs and private forecasters' predictions that we will have to take care of. Nothing ensures that private forecasts are good or close to optimal forecasts. If private forecasts were poor, they would be of little use as benchmarks.

To overcome this potential problem and complement our analysis, we also proceed to compare the AIFs with simple forecasts from time series models. Following a preliminary evaluation process, we decided to use an ARMA (p,q) model estimated with rolling windows of 30 observations and automatic p and q parameter selection according to Akaike's criteria. The estimate is made imposing the restriction that long-term inflation exactly matches the Central Bank of Chile's inflation target (3%). Of all the ARMA methods we explored, the one described above presented the lowest MSPE out-of-sample at most of the horizons considered.

4.2 Bias and weak efficiency

We also evaluate two properties associated with an optimal prediction error: zero bias and weak efficiency. While the evaluation of a zero bias is easily accommodated into Giacomini and White's (2006) framework, measuring it simply as the expected value of prediction errors, the efficiency test we use was originally introduced by Mincer and Zarrowitz (1969) and it is based on a simple regression between the prediction error and the predictor itself. The null hypothesis is that the predictor has no statistically significant coefficient associated. If the null hypothesis is rejected we can conclude that the prediction has not been efficient in the sense of using adequately the available information.

4.3 Encompassing

We also want to evaluate whether private forecasters' predictions can contribute to improve the AIFs. This is usually done using forecast encompassing tests. Granger and Ramanathan (1984) suggest to test encompassing by regressing a forecast's errors over a constant term and an alternative forecast. If this alternative forecast is able to explain the original forecast errors then we can say that this initial forecasting method does not encompass the alternative method.

4.4 Dealing with different information sets

The timeline in computing forecasts is important to assess whether the differences in predictive accuracy between two series of forecasts originate simply in the different timing in which these forecasts are generated. Casual talks with private analysts reveal that their forecasts are often revised in the light of news on key variables such as oil prices and exchange rates. This would imply that prediction errors should tend to be lower when the forecasts are made further into the month and forecasters have more relevant information.

The timeline in the construction of forecasts indicates that ARMA predictions are the ones using least information since they are only based on past effective inflation. On the other hand, the SPF forecasts are usually built during the first week of the month so they could potentially use more information than that used for ARMA predictions. At the closure of the AIFs, the SPF information is already known so the AIFs potentially count on more information than the SPF. Later the Consensus Forecast survey is published and finally, towards the end of the month, results of Bloomberg's survey are published. Consequently, if there were no significant differences in the analysts' ability, we could naturally expect that - on average - Bloomberg's predictions would be at least as accurate as all the others, that Consensus Forecast's predictions would be at least as accurate as those of the SPF, the AIF and the ARMA method, that the AIF forecasts were at least as accurate as those of the SPF and the ARMA, that the SPF's predictions were at least as accurate as the ARMA, and that the ARMA were the weakest of all. Any deviation from the foregoing pattern would have to indicate that one agent had greater predictive ability than another.

This way of reaching conclusions would indicate, for example, that a lower MSPE in the AIFs than in the SPF forecasts is to be expected and should not necessarily be considered as evidence of better predictive ability of the forecasters in charge of the AIF. From now on we will refer to this expected order of predictive accuracy as hypothesis 1 (H1). This hypothesis is summarized as follows:

H1: Bloomberg>>Consensus>>AIF>>SPF>>ARMA

in which the double inequality operator should be read from left to right as "at least as accurate as". We will be rigorous in detecting deviations from H1.

To complement the search for deviations from H1 we want to point out another pattern that must be satisfied by optimal forecasts that only differ in the moment at which they are built. Let us suppose that we want to test the hypothesis that two forecasters have exactly the same ability to predict inflation. And let us suppose that forecaster 1 builds his predictions in an optimal way using information up to time t_0, while forecaster 2 also builds his predictions in an optimal way but with information available up to time $t_1>t_0$.

The optimal h-step ahead prediction error under quadratic loss is defined as the difference between the original series in predicting Y_{t+h} and the optimal predictor $Y^f_t(h)$. We express this error as $e_t(h)$ and write it as:

$$e_t(h) = Y_{t+h} - Y_t^f(h)$$
$$e_t(h) = \sum_{i=0}^{\infty}\psi_i\varepsilon_{t+h-i} - \sum_{i=h}^{\infty}\psi_i\varepsilon_{t+h-i}$$
$$e_t(h) = \sum_{i=0}^{h-1}\psi_i\varepsilon_{t+h-i}$$

in which we have assumed the following Wold representation for the original series:

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

Let us also assume that $t_1 = t_0 + k$, thus

$$e_{t0}(h+k) = \sum_{i=0}^{h+k-1} \psi_i \varepsilon_{t0+h+k-i}$$

$$e_{t1}(h) = \sum_{i=0}^{h-1} \psi_i \varepsilon_{t1+h-i}$$

We note that

$$e_{t0}(h+k) = \sum_{i=0}^{h+k-1} \psi_i \varepsilon_{t0+h+k-i}$$

$$e_{t0}(h+k) = \sum_{i=0}^{h-1} \psi_i \varepsilon_{t0+h+k-i} + \sum_{i=h}^{h-1+k} \psi_i \varepsilon_{t0+h+k-i}$$

$$e_{t0}(h+k) = \sum_{i=0}^{h-1} \psi_i \varepsilon_{t1+h-i} + \sum_{i=h}^{h-1+k} \psi_i \varepsilon_{t0+h+k-i}$$

$$e_{t0}(h+k) = e_{t1}(h) + \sum_{i=h}^{h-1+k} \psi_i \varepsilon_{t0+h+k-i}$$

Hence, if the only difference between the two forecasters is that one makes his forecast later than the other (in the case of our example, a difference of k periods), we see that for a fixed prediction horizon $T = t_1 + h = t_0 + h + k$, the MSPE differential between the two forecasters is:

$$E\left(e_{t0}(h+k)\right)^2 - E(e_{t1}(h))^2 = \sum_{i=h}^{h+k-1} \psi^2{}_i E(\varepsilon_{t0+h+k-i})^2 = \sigma_\varepsilon^2 \sum_{i=h}^{h+k-1} \psi^2{}_i$$

and the MSPE ratio between them looks as follows

$$\frac{E\left(e_{t0}(h+k)\right)^2}{E(e_{t1}(h))^2}=1+\frac{\sigma_\varepsilon^2\sum_{i=h}^{h+k-1}\psi^2{}_i}{E(e_{t1}(h))^2}$$

$$\frac{E\left(e_{t0}(h+k)\right)^2}{E(e_{t1}(h))^2}=1+\frac{\sigma_\varepsilon^2\sum_{i=h}^{h+k-1}\psi^2{}_i}{E(\sum_{i=0}^{h-1}\psi_i\varepsilon_{t1+h-i})^2}$$

$$\frac{E\left(e_{t0}(h+k)\right)^2}{E(e_{t1}(h))^2}=1+\frac{\sigma_\varepsilon^2\sum_{i=h}^{h+k-1}\psi_i^2}{\sigma_\varepsilon^2\sum_{i=0}^{h-1}\psi_i^2}$$

$$\frac{E\left(e_{t0}(h+k)\right)^2}{E(e_{t1}(h))^2}=1+\frac{\sum_{i=h}^{h+k-1}\psi_i^2}{\sum_{i=0}^{h-1}\psi_i^2}\xrightarrow{h\to\infty}1^+$$

Thus, when the prediction horizon lengthens indefinitely (that is h grows with fixed k) then

$$\frac{E\left(e_{t0}(h+k)\right)^2}{E(e_{t1}(h))^2}=1+\frac{\sum_{i=h}^{h+k-1}\psi_i^2}{\sum_{i=0}^{h-1}\psi_i^2}\xrightarrow{h\to\infty}1^+$$

This is because the right hand side numerator goes to zero as h goes to infinity, since the infinite sum of Wold´s theorem squared coefficients is finite, whilst the denominator is growing, since, while h grows, positive terms are added.

Accordingly, under the assumption of optimality and if the only difference between the two forecasters were the moment at which they make their prediction for a certain horizon, the MSPE ratio at different horizons would generate a path towards 1. If the MSPE of the forecaster who makes predictions with less available information is placed in the numerator, this path will have a decreasing trend, and otherwise it will have an increasing trend.

This pattern constitutes our second working hypothesis, which we will call H2. We explore by visual inspection of the MSPE ratios whether the predictive accuracy differences can be justified only on the grounds of the predictions being made at different points in time.

5 Comparison with private analysts and simple time series models

In this section we compare the AIFs with private analysts' forecasts. In the first subsection we rank their predictive accuracy on the basis of MSPE comparisons in terms of H1. In the second subsection, we evaluate the differences in terms of H2. In subsection 3 we use inference to determine the statistical significance of the differences we found in the available sample.

5.1 Descriptive analysis with regard to H1

Table 1 shows the relative position of the AIFs in terms of MSPE, which is our chosen measure of predictive accuracy, over the 6 horizons considered [8].

Compared with the ARMA forecasts, the AIFs have lower MSPE for 1-month ahead to 5-month ahead horizons but the ARMA forecasts 6-month ahead are more accurate. Compared with the SPF's results, the AIFs show the least MSPE in 1-month ahead inflation forecasts but the performance is slightly poorer at 3 months ahead with the AIFs falling to third place in the ranking of predictive accuracy. Compared with the Consensus Forecast's results, the AIFs are better at 1-month and 2-month ahead horizons but the Consensus forecasts are more accurate at longer horizons of 3, 4, 5 and 6 months ahead.

The comparison with Bloomberg's forecasts is not so benevolent as with the other competing forecasts. The AIFs come 6th out of 7 in the 1-month ahead forecast which is the only horizon considered for the Bloomberg survey.

8 In the case of the SPF and Bloomberg's survey, we have considered the mean and the median of the forecasts as two more analysts.

In summary, according to the evidence thus far, the AIFs behave in line with our working hypothesis H1, with the two following important exceptions:

1. When forecasting 1 month and 2 months ahead, the AIFs behave better than expected. This is because they are not only more accurate than forecasts built with less available information, but also they are more accurate than Consensus forecasts and those from 5 analysts considered in the Bloomberg´s survey. This is an evident deviation from our H1 hypothesis, because Consensus forecasts and those from the Bloomberg´s survey are built counting on information not available when the AIFs are constructed.

2. At horizons of 3 months ahead, the AIFs forecasts are outperformed by two SPF analysts. This is surprising given that the AIFs consider more information than that available when the SPF is carried out. At 6-month ahead horizons, we see that the ARMA forecasts rank higher than the AIFs. This evidence suggests that the AIFs' relative performance deteriorates as the prediction horizon lengthens[9].

An interesting point to note is that the two SPF forecasts that beat the AIFs at 3-month horizons come from individual analysts. That is, they do not correspond to either the mean or the median of the analysts. In fact, if we look at Charts A1 and A2 in the Annex, we see that the simple average and the median only fare moderately well in the predictive accuracy ranking.

Chart A1 shows the ratio between the MSPE of each 1-month ahead prediction of the SPF and the MSPE of the AIFs. Chart A2 is analogous but using 3-month ahead forecasts. These charts show that the simple average of the individual forecasts behaves very much the same as the median and that both aggregates show moderate performance in terms of predictive accuracy. At 1-month ahead horizons, the mean and the median rank 10th and 9th among all the SPF respondents. At 3 months ahead they rank worse: 13th and 14th respectively.

---

9 Consensus' forecasts are also more precise than those of the AIF at 3, 4, 5 and 6 months ahead, but this is in line with our H1 hypothesis and therefore this evidence does not allow any conclusion about the AIF forecasters' predictive ability in comparison with that of the Consensus forecasters. We return to this point later.

This result is curious because of the number of papers indicating that predictive accuracy can be improved by an adequate combination of forecasts and that the combined forecast can even be more accurate that the best individual forecasts. Additionally, this literature sets forth the so-called "combination puzzle" which claims that simple combination methods in general behave better than optimal or more complex methods. If this were really so, then the empirical evidence presented here would indicate either that the average is not an adequate way of combining or that no combination can outperform the best individual predictions. Both assertions are to some extent in contradiction with the traditional forecasting combination literature.

Table 1

Descriptive Analysis of Predictive Evaluation

Horizon	ARMA	SPF	Consensus	Bloomberg
1 month	1	1	1	6
2 months	1	na	1	na
3 months	1	3	2	na
4 months	1	na	2	na
5 months	1	na	2	na
6 months	2	na	2	na
N	**2**	**37/36**	**2**	**7**

1. Each cell displays the rank of the Central Bank of Chile Auxiliary Inflation Forecasts (AIF) in terms of MSPE compared to different forecasts at different horizons.

2. N stands for the number of competing forecasts. For instance, N=2 when the AIF are compared to ARMA forecasts and N=7 when the AIF are compared to Consensus Forecasts.

3. N= 37/36 means 37 forecasts were considered for 1 month ahead comparisons and 36 were considered for 3 step ahead comparisons.

4. na stands for not available.

Source: Author's calculations.

5.2 Descriptive analysis with regard to H2

Hypothesis 2 claims that if the only difference between two forecasters were the moment at which they make their forecast, then the MSPE ratio at different horizons would generate a path towards 1.

We have two sources of forecasts at various horizons: Consensus and ARMA forecasts. Figure 3 shows the ratio of MSPE between Consensus and the AIFs at various horizons. Similarly, Figure 4 shows the ratio of MSPE between ARMA forecasts and the AIFs at the six horizons considered.

Figure 3: Ratio of MSPE between Consensus and AIF

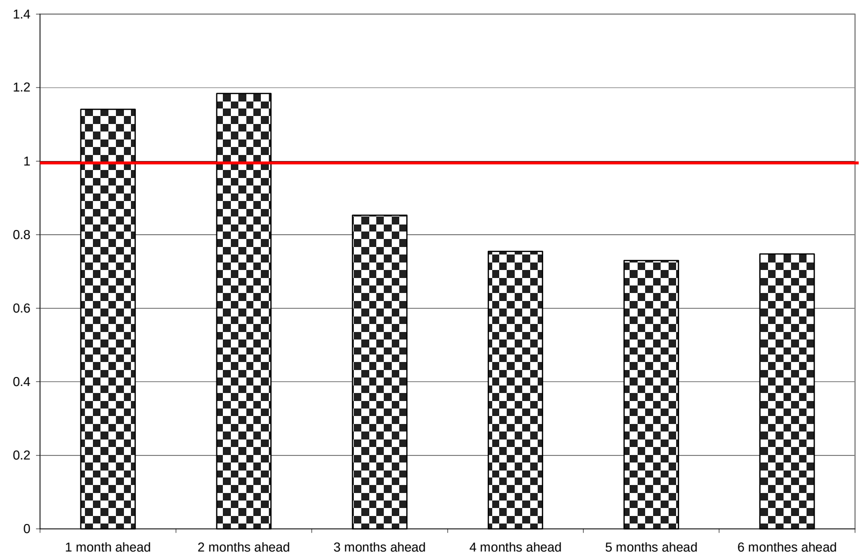


Figure 4: Ratio of MSPE between ARMA and AIF

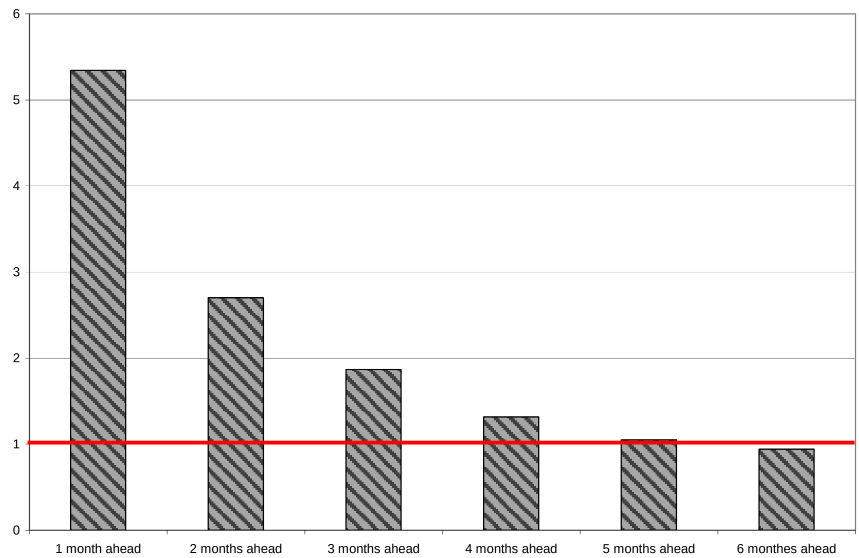


Figure 4 shows a decreasing pattern of the ratio of MSPEs that is almost 100% compatible with H2, whereas Figure 3 shows a pattern that is difficult to reconcile with this hypothesis. According to H2, the ratio between the MSPE of the ARMA and AIF predictions should start being greater than 1 to decrease and converge asymptotically to 1. The only exception to this behavior is observed in the 6-month ahead forecast where the MSPE ratio drops slightly below 1. This small exception, however, does not report any additional information to that already shown in Table 1, first column. Therefore, Figure 4 shows a pattern that in general is consistent with H2, which means that the greater accuracy of the AIFs over ARMA predictions is possible due to differences in the information sets at the moment of prediction.

The interpretation of Figure 3 is more complex but also more informative. According to our hypothesis H2, we expect the MSPE ratio to start from a value lower than 1 and grow asymptotically towards 1. Figure 3 shows almost exactly the opposite behavior, starting at values greater than 1 and decreasing to values well below 1. We have already shown in Table 1 that at short horizons the AIFs are more accurate than Consensus forecasts and that this result is reverted at horizons of over two months.

Nevertheless, Figure 3 provides additional information: the fact that MSPE ratios stagnate at values distant from 1 suggests that the differences in forecasting accuracy at longer horizons might not be explained by a possible difference in the information sets available when the forecasts are initially made. This evidence suggests that the AIFs are systematically less accurate than the Consensus forecasts at horizons over two months, irrespective of the potential differences in initial information. It is worth mentioning, however, that this last conclusion is made under the assumption that a forecasting horizon of 6 months is actually a very long horizon, which is, of course, a strong assumption.

5.3 Statistical inference

Table 2 summarises the information of Giacomini and White's (2006) predictive ability tests to compare the performance of the AIFs and the other data sources. This table highlights the statistically significant differences in MSPE at a 10% significance level.

Table 2

Statistical inference

	Predictive Horizon						
Restricted ARMA	**1**	**2**	**3**	**4**	**5**	**6**	
	1 / 1	1 / 1	0 / 1	0 / 1	0 / 1	0 / 1	[AIF is more accurate]
	0 / 1	0 / 1	0 / 1	0 / 1	0 / 1	0 / 1	[AIF is less accurate]
	Predictive Horizon						
Survey of Professional Forecasters	**1**	**2**	**3**	**4**	**5**	**6**	
	30 / 36	-	21 / 35	-	-	-	[AIF is more accurate]
	0 / 36	-	0 / 35	-	-	-	[AIF is less accurate]
	Predictive Horizon						
Consensus Forecast	**1**	**2**	**3**	**4**	**5**	**6**	
	0 / 1	0 / 1	0 / 1	0 / 1	0 / 1	0 / 1	[AIF is more accurate]
	0 / 1	0 / 1	0 / 1	0 / 1	0 / 1	1 / 1	[AIF is less accurate]
	Predictive Horizon						
Bloomberg	**1**	**2**	**3**	**4**	**5**	**6**	
	0 / 6	-	-	-	-	-	[AIF is more accurate]
	2 /6	-	-	-	-	-	[AIF is less accurate]

1. Each cell displays findings of statistically significant differences in MSPE between the Central Bank of Chile Auxiliary Inflation Forecasts (AIF) and different sources of forecasts.
2.For instance, the expression 1/1 in the upper row on each panel, means that the AIC is more accurate than the respective competing forecast. The same expression 1/1 located in the lower row in the same panel means that the AIC is less accurate than the respective competing forecast.
3. Displayed results are evaluated at the 10% significance level.

Source: Author's calculations.

Interestingly, Table 2 shows that all the results in conflict with H1 shown in Table 1 are not statistically significant. In other words, Table 2 results are completely coherent with H1. The lower MSPE of the ARMA forecasts compared to the AIF forecasts at 6-month ahead horizon is not statistically significant. The same occurs with the two SPF forecasters who beat the AIF in terms of MSPE at 3 month horizons. Similarly, the AIFs no longer beat any Bloomberg forecaster nor the Consensus forecast at 1 and 2 month horizons. However, these results should be taken cautiously since the samples only amount to at most 84 observations (January 2001-December 2007) which could affect the size and power of these predictive ability tests.

To sum up, our results show that the relative AIF performance is as expected, given the information available at the moment of prediction. Some exceptions suggest that the AIFs appear to perform better than expected at horizons of up to two months ahead and a little worse than expected at longer horizons, but Table 2 shows that some of these exceptions have no statistical significance.

6 Bias and Efficiency

6.1 Bias

Forecasts are unbiased if the expected prediction error is equal to zero. This is a property that in theory should satisfy an optimal prediction error under quadratic loss. In this article we will conclude that a forecast is biased if the hypothesis Ho: $\alpha = 0$ is rejected in the following regression:

$$(Y_t - Y^f_t) = \alpha + \varepsilon_t$$

in which Y_t denotes the series to be forecast, Y^f_t corresponds to the forecast of Y_t and ε_t represents a random shock. Results of the bias analysis are shown in Table 3. In general we see that we cannot reject the hypothesis of no bias for the great majority of analysts and methods considered as benchmarks for the AIFs, although the statistics' signs indicate that all the analysts and methods in general tend to underestimate inflation.

In the particular case of the AIFs, the no bias hypothesis is rejected for 4, 5 and 6 month ahead forecasts. We also notice a growing tendency to underestimate inflation as the horizon lengthens, reaching over 50 basis points in 6-month ahead forecasts.

Table 3

Forecast Bias

PANEL A: BIAS						
	1	**2**	**3**	**4**	**5**	**6**
Central Bank of Chile AIF	0.04	0.07	0.20	0.36*	0.47*	0.52*
Restricted ARMA	0.16	0.21	0.27	0.28	0.30	0.30
Survey of Professional Forecasters	0.03	-	0.10	-	-	-
Consensus Forecast	0.00	0.05	0.10	0.15	0.18	0.20
Bloomberg	0.02	-	-	-	-	-
PANEL B: Number of Unbiased Forecasters						
	1	**2**	**3**	**4**	**5**	**6**
Survey of Professional Forecasters	33 / 36	-	34 / 35	-	-	-
Bloomberg	6 / 6	-	-	-	-	-

1. - *Rejection at 10%, **Rejection at 5%, ***Rejection at 1%
2. – Results displayed in panel A for the Survey of Professional Forecasters and the survey carried out by Bloomberg are calculated for the median forecaster.

Source: Author's calculations.

It is outstanding that none of the six measures from Bloomberg are statistically biased and that from the SPF only three measures at 1 month and one measure at 3 months have a positive and statistically significant bias.

6.2 Efficiency

Efficient data processing requires that prediction errors have no correlation with past available information. This property is also a distinctive feature of an optimal prediction error under quadratic loss. In this article we use a weak efficiency concept that restricts the available set of information to a constant term and to the actual prediction value. The estimated regression is as follows:

$$(Y_t - Y_t^f) = \beta_1 + \beta_2 Y_t^f + \upsilon_t$$

In which Y_t denotes the series to be forecast, Y^f_t corresponds to the prediction of Y_t and ν_t represents a random shock. The prediction is considered efficient if the null hypothesis $\beta_2 = 0$ cannot be rejected.

As shown in Table 4, for most cases and analysts, the hypothesis that the predictions are efficient cannot be rejected. The only exceptions are one of the six Bloomberg forecasts, a few SPF forecasts, ARMA forecasts at 5 and 6 month horizons and the AIFs at the 6 month horizon[10].

Table 4

Forecast efficiency

PANEL A: WEAK EFFICIENCY						
	1	**2**	**3**	**4**	**5**	**6**
Central Bank of Chile AIF	0.03	0.03	-0.02	-0.10	-0.23	-0.39*
Restricted ARMA	0.05	0.16	0.24	-0.10	-0.59**	-0.94***
Survey of Professional Forecasters	0.03	-	-0.04	-	-	-
Consensus Forecast	0.01	0.02	-0.04	-0.07	-0.12	-0.20
Bloomberg	0.03	-	-	-	-	-
PANEL B: RATE OF WEAKLY EFFICIENT FORECASTERS						
	1	**2**	**3**	**4**	**5**	**6**
Survey of Professional Forecasters	32 / 36	-	26 / 35	-	-	-
Bloomberg	5 / 6	-	-	-	-	-

1. - *Rejection at 10%, **Rejection at 5%, ***Rejection at 1%
2. – Results displayed in panel A for the Survey of Professional Forecasters and the survey carried out by Bloomberg are calculated for the median forecaster

Source: Author's calculations.

7 Encompassing analysis

Forecast A is said to encompass forecast B if the information offered by B is not useful for diminishing the prediction error of A. To test whether forecast A encompasses forecast B we run the following regression:

$$(Y_t - Y_t^A) = \lambda_1 + \lambda_2 Y_t^B + u_t$$

[10] We must point out that the fact of using a rolling window estimate of 30 observations to obtain the ARMA forecasts considerably reduces the number of observations available for building statistics. This means that the comparisons of bias and efficiency for the ARMA forecasts are difficult to compare with the rest of the methods.

where Y_t denotes the series to be predicted, Y^A_t coresponds to forecaster A's prediction of Y_t, Y^B_t corresponds to forecaster B's prediction of Y_t and u_t represents a random shock. We conclude that A encompasses B if $H_0 : \lambda_2 = 0$ cannot be rejected.

In this case we are testing whether the AIFs encompass the forecasts of methods available when the AIFs were made, that is, whether the forecasts made prior to the AIF have any useful information for reducing the AIFs' prediction errors. Table 6 shows the results.

The evidence shows that the AIFs encompass the median of the SPF only for 3-months ahead forecasts. Furthermore the AIFs encompass an important number of individual forecasts of that survey when 1-month ahead predictions are considered. For predictions 3-months ahead the AIFs encompass most of the individual forecasts of the survey. Nevertheless, this analysis reveals that there is information contained in a few SPF forecasts that could be useful for reducing AIF prediction errors.

Table 6

Encompassing Analysis

PANEL A: ENCOMPASSING TESTS						
	1	**2**	**3**	**4**	**5**	**6**
Restricted ARMA	0.03	0.03	-0.04	-0.17	-0.31	-0.39
Survey of Professional Forecasters	0.04*	-	0.00	-	-	-
PANEL B: RATE OF EMCOMPASSED FORECASTERS						
	1	**2**	**3**	**4**	**5**	**6**
Survey of Professional Forecasters	24 / 36	-	33 / 35	-	-	-

1. - *Rejection at 10%, **Rejection at 5%, ***Rejection at 1%
2. – Results displayed in panel A for the Survey of Professional Frecasters are calculated for the median forecaster.

Source: Author's calculations.

8. Conclusions

In this article we evaluate the Central Bank of Chile's Auxiliary Inflation Forecasts (AIF), comparing them with forecasts from simple time series models and forecasts from three additional sources: Survey of Professional Forecasters, Consensus Forecast and Bloomberg. We also evaluate properties of bias, weak efficiency and encompassing of the forecasts.

A key aspect to consider is the forecast construction timeline. It can be reasonably expected that forecasts made with more information will be more accurate than forecasts based on less information. It is also reasonable to expect that this difference will be reduced as the forecast horizon lengthens. In this sense, our results show that the AIFs perform relatively as expected given the information available at the moment they were made, although some notable exceptions were detected:

1. The AIFs predictive accuracy at horizons of less than 3 months is better than expected and they are even more accurate than some forecasts probably made after them. Moreover, at these horizons the AIF survives both tests of bias and weak efficiency.

2. The AIFs predictive accuracy at horizons of more than 2 months is not as good as expected. In fact, AIF are even less accurate than some forecasts probably made before them. We might expect that AIF performance would not compare well with Consensus forecasts that are made after the AIF forecasts, but our exercises suggest that the differences apparently cannot be justified in terms of the available information at the moment of prediction. Moreover, at some horizons longer than two months the AIFs do not survive the tests of bias and weak efficiency.

Another finding from the encompassing analysis suggests that the AIFs could benefit from using the information contained in a few SPF forecasts.

Finally we would like to point out an interesting unexpected collateral finding. Contrary to what is shown in much of the forecast combination literature, the simple mean of the set of forecasts coming from the Central Bank of Chile's SPF has a quite moderate performance compared to the individual forecasts. This result is valuable because it reveals concrete empirical evidence against what is known as the "combination puzzle".

Annex

Chart A1
MSPE ratio between SPF and AIF
1-month ahead forecasts

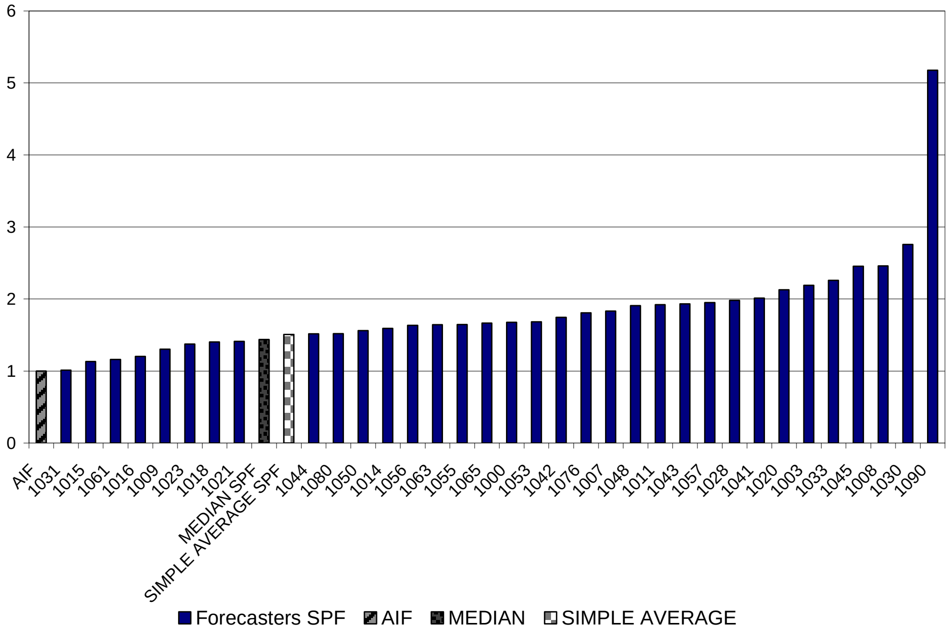
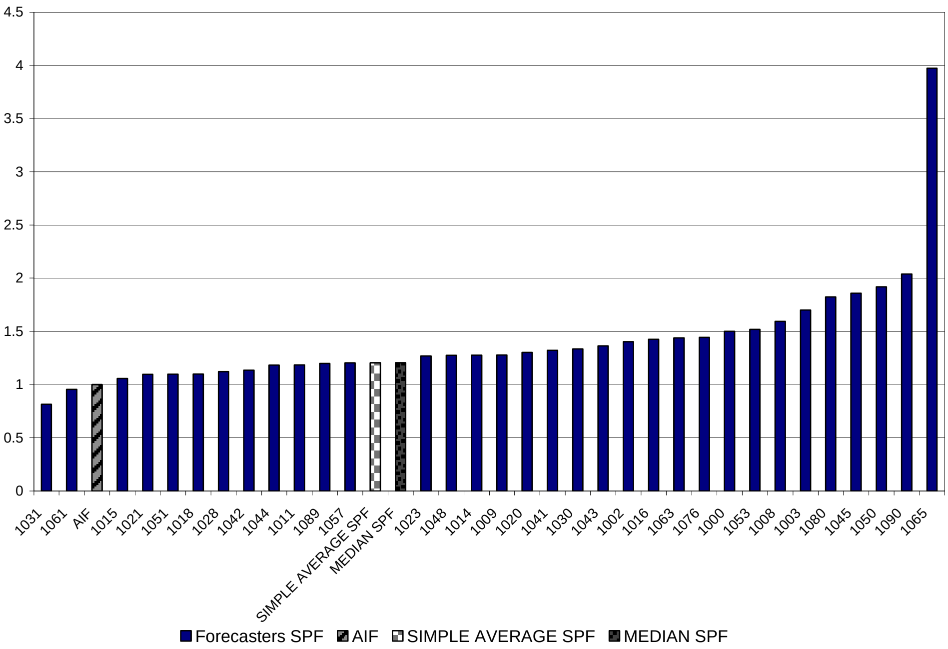


Chart A2
MSPE ratio between SPF and AIF
3-month ahead forecasts



References

Albagli, E., Contreras, G., García, P., Magendzo, I. and R. Valdés (2003): “Errores de Proyección en Perspectiva,” Working Paper No. 199, Central Bank of Chile

Ang, A., Bekaertb G. and M. Wei (2007): “Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?”, *Journal of Monetary Economics* 54(4): 1163-1212.

Andersson, M., Karlsson, G. and J. Svensson (2007): "The Riksbank's Forecasting Performance", Riksbank Research Paper No. 218.

Bentancor A. and P. Pincheira (2008): “Prediciendo los Errores de Proyección de Inflación en Chile”. Working Paper Nº 477 Central Bank of Chile.

Bowles, C., Friz, R., Genre, V., Geoff, K., Aidan, M. and R. Tuomas (2007): “The ECB Survey of Professional Forecasters (SPF) - A Review After Eight Years' Experience,” European Central Bank Occasional Paper No. 59.

Capistrán C. and G. López-Moctezuma (2008): “Las Expectativas Macroeconómicas de los Especialistas: Una Evaluación de Pronósticos de Corto Plazo en México”. Working Paper Nº 2008-11 Bank of México.

Chumacero, R.A. (2001): “Empirical Analysis of Systematic Errors in Chilean GDP Forecast,” *Journal of Forecasting* 20(1): 37-45.

Croushore, D. (1998): "Evaluating inflation forecasts," Working Papers 98-14, Federal Reserve Bank of Philadelphia.

Diebold F. and R. Mariano (1995): “Comparing Predictive Accuracy,” *Journal of Business and Economic Statistics* 13: 253-263.

Granger C. and R. Ramanathan (1984): “Improved Methods of Combining Forecast,” *Journal of Forecasting* 3: 197-204.

Giacomini, R. and H. White (2006): "Tests of Conditional Predictive Ability," Econometrica. Vol 74 N° 6.

Leitch G. and E. Tanner (1991): “Economic Forecast Evaluation: Profits versus Conventional Error Measures,” *American Economic Review* 81: 580-590.

Loungani, P. (2001): “How Accurate are Private Sector Forecast? Cross-Country Evidence from Consensus Forecast of Output Growth,” *International Journal of Forecasting* 17(3): 419- 432.

McCulloch R. and P. Rossi (1990): “Posterior, Predictive and Utility-Based Approaches to Testing the Arbitrage Pricing Theory,” *Journal of Financial Economics* 28:7-38.

Mincer J. and V. Zarnowitz (1969): “The Evaluation of Economic Forecast.” en J. Mince (Ed.): *Economic Forecast and Expectations*. New York NBER.

Newey, W. K. and K D West (1994): “Automatic Lag Selection in Covariance Matrix Estimation,” *Review of Economic Studies* 61(4):631-53.

Newey, W.K. and K.D. West (1987): “A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica* 55: 703-08.

Oller, L. and B. Barot (2000): “The Accuracy of European Growth and Inflation Forecast,” *International Journal of Forecasting* 16(3): 293-315.

Smith J. and K.F. Wallis (2008): “A Simple Explanation of the Forecast Combination Puzzle”, manuscript Department of Economics, University of Warwick, September 2008.

Stock, J.H. and Watson, M.W. (2004): ‘Combination forecasts of output growth in a seven country data set’, *Journal of Forecasting*, Vol. 23, pp. 405-430.

West K, H. Edison and D. Cho (1993): "A Utility-Based Comparison of Some Models of Exchange Rate Volatility," *Journal of International Economics* 35: 23-45.

West K. (1996): "Asymptotic Inference About Predictive Ability". Econometrica 64: 1067-1084.

Documentos de Trabajo Banco Central de Chile

Working Papers Central Bank of Chile

NÚMEROS ANTERIORES

PAST ISSUES

La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: www.bcentral.cl/esp/estpub/estudios/dtbc. Existe la posibilidad de solicitar una copia impresa con un costo de $500 si es dentro de Chile y US$12 si es para fuera de Chile. Las solicitudes se pueden hacer por fax: (56-2) 6702231 o a través de correo electrónico: bcch@bcentral.cl.

Working Papers in PDF format can be downloaded free of charge from: www.bcentral.cl/eng/stdpub/studies/workingpaper. Printed versions can be ordered individually for US$12 per copy (for orders inside Chile the charge is Ch$500.) Orders can be placed by fax: (56-2) 6702231 or e-mail: bcch@bcentral.cl.

DTBC – 673 Julio 2012
Tasa Máxima Convencional y Oferta de Créditos
Rodrigo Alfaro, Andrés Sagner y Camilo Vio

DTBC – 672 Julio 2012
Peg, Downward Wage Rigidity, and Unemployment: The Role of Financial Structure
Stephanie Schmitt- Grohé y Martín Uribe

DTBC – 671 Julio 2012
Adapting Macroprudential Policies to Global Liquidity Conditions
Hyun Song Shin

DTBC – 670 Julio 2012
An Anatomy of Credit Booms and their Demise
Enrique Mendoza y Marco Terrones

DTBC – 669 Julio 2012
Forecasting Inflation with a Random Walk
Pablo Pincheira y Carlos Medel

DTBC – 668 Junio 2012
On the International Transmission of Shocks: Micro Evidence From Mutual Fund Portfolios
Claudio Raddatz y Sergio L. Schmukler

DTBC – 667 Mayo 2012
Heterogeneous Inflation Expectations, Learning and Market Outcomes
Carlos Madeira y Basit Zafar

DTBC – 666 Marzo 2012
Financial Development, Exporting and Firm Heterogeneity in Chile
Roberto Alvarez y Ricardo López

DTBC – 665 Marzo 2012
Determinantes e Impacto de Episodios de Reversión Abrupta de Flujos de Capitales: ¿Es Distinto un *Sudden Stop* de un *Sudden Flight*?
Gabriela Contreras, Alfredo Pistelli y Mariel Siravegna

DTBC – 664 Febrero 2012
Rational Inattention, Multi-Product Firms and the Neutrality of Money
Ernesto Pastén

DTBC – 663 Febrero 2012
Non – Ricardian Aspects of the Fiscal Policy in Chile
Luis Felipe Céspedes, Jorge Fornero y Jordi Galí

DTBC – 662 Febrero 2012
Cubrir o no Cubrir: ¿Ese es el Dilema?
Rodrigo Alfaro y Natán Goldberger

DTBC – 661 Febrero 2012
Are Forecast Combinations Efficient?
Pablo Pincheira

DTBC – 660 Enero 2012
Combinación de Proyecciones para el Precio del Petróleo: Aplicación y Evaluación de Metodologías.
Ercio Muñoz, Miguel Ricaurte y Mariel Siravegna

BANCO CENTRAL
DE CHILE