What Do Forecasters Say?
Should We Care? [†]

Rómulo A. Chumacero [‡]

ABSTRACT

This paper presents a statistical comparison between the actual and predicted evolution of the Chilean GDP for the period 1986-1998 made by several forecasters. We show that the forecasters systematically underestimate the true growth rate of the economy. The magnitude of this bias tends to be correlated with the phase of the business cycle. We briefly evaluate if there is any reason why we should care about what they say.

Keywords: Forecasting, Business Cycles.

JEL Classification: C53, E32, E37, E52.

Santiago, February 1999

[†] I would like to thank Rodrigo Fuentes, Osvaldo Larrañaga, Ricardo Paredes and José Miguel Sánchez for helpful comments. Carlos Oyarzún and Patricia Toledo provided able research assistance. The usual disclaimer applies.

[‡] Department of Economics. University of Chile. E-mail: rchumace@decon.facea.uchile.cl.

1 Introduction

The Chilean economy displays one of the highest average growth rates in the Western Hemisphere. In fact, the average annual growth rate of the GDP between 1986 and 1997 is of approximately 7.7%.

Accompanying this spectacular performance, there has been a proliferation of projections of the evolution of the economy by different sources. In this work we evaluate these projections by comparing them with the actual evolution of the rate of growth of the GDP for the period comprehended between 1986 and 1997. We gathered 857 forecasts made by different sources and published in the financial newspaper *Estrategia* during the same period. These forecasts correspond to projections made by 43 individual forecasters, 28 organizations, 16 commercial banks, 23 private companies, 7 insurance companies, and 11 pension funds.

The paper is organized as follows: Section 2 presents a comparison between the effective evolution of the growth rate of the GDP and the projections made by the forecasters. Section 3 shows how forecasts based on very simple models are able to outperform the forecasters. Finally, Section 4 concludes.

2 How well do forecasters do?

This section analyzes the statistical properties of the forecast errors made by several forecasters during the period comprehended between 1986 and 1997. Our objective is not to provide an explanation of why they incurred in these errors but to stress some of the empirical regularities that can be associated with them.

2.1 How are they distributed?

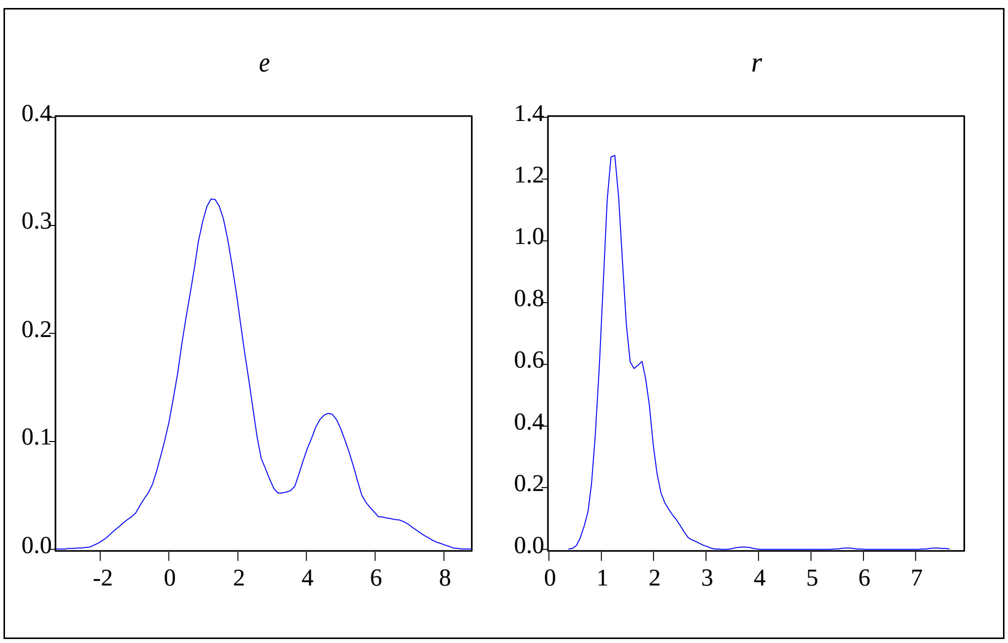
In this paper, we define as absolute (relative) forecast errors ($e_{i,t}$ and $r_{i,t}$ respectively) made by forecaster i in period t as the difference (ratio) among the effective value of the growth rate of GDP in period t (g_t) and the forecast made by the forecaster i for that period ($f_{i,t}$). That is:

$$e_{i,t} = g_t - f_{i,t}; \qquad r_{i,t} = \frac{g_t}{f_{i,t}}$$

Thus, according to our definition, an underestimation of the growth rate would lead to a positive value for *e* and a value of *r* greater than one. Conversely, an overestimation would lead to negative values for *e* and values of *r* smaller than one.

Let us begin then defining some of the most important properties that have these variables. For that purpose, we build a vector of absolute and relative errors to evaluate their unconditional distribution. As mentioned previously, 857 observations were available. Figure 1 shows the non-parametric estimators of the unconditional distributions of both series, while Table 1 displays some of their descriptive statistics.

Figure 1
Unconditional Densities of *e* and *r*



Note: The unconditional densities were estimated using the Epanechnikov kernel and the bandwidth selection proposed by Silverman (1986).

As evidenced, the unconditional distributions of both variables show strong departures of normality. Both distributions are bimodal and asymmetric. Notice in both cases the important bias towards underestimation. That is *e* is biased towards positive values and *r* is biased towards values exceeding 1. In fact, simple tests show that the null of unbiased forecast errors is strongly rejected in both cases.[1] More importantly, on average, the forecasters underestimated the growth rate of GDP by more than 2 points.

Table 1
Descriptive Statistics of *e* and *r*

Statistic	*e*		*r*	
	Value	*p*-value	Value	*p*-value
Mean	2.165	0.000	1.458	0.000
Median	1.600	0.000	1.325	0.000
Standard Deviation	1.892		0.477	
CV	0.874		0.327	
S	0.727	0.000	3.751	0.000
K	2.821	0.269	38.171	0.000
JB	76.651	0.000	46180.890	0.000

Note: Mean = The *p*-values correspond to the nulls that the mean of *e* is 0 and the one of *r* is 1. Medium = The *p*-values correspond to the null that the median of *e* is 0 and the one of *r* is 1. CV = Coefficient of Variation. *S* = Skewness. The *p*-value corresponds to the null that *S* is 0. *K* = Kurtosis. The *p*-value corresponds to the null that *K* is 3. *JB* = Jarque and Bera Normality test. The *p*-value corresponds to the null of $S = 0$ and $K = 3$.

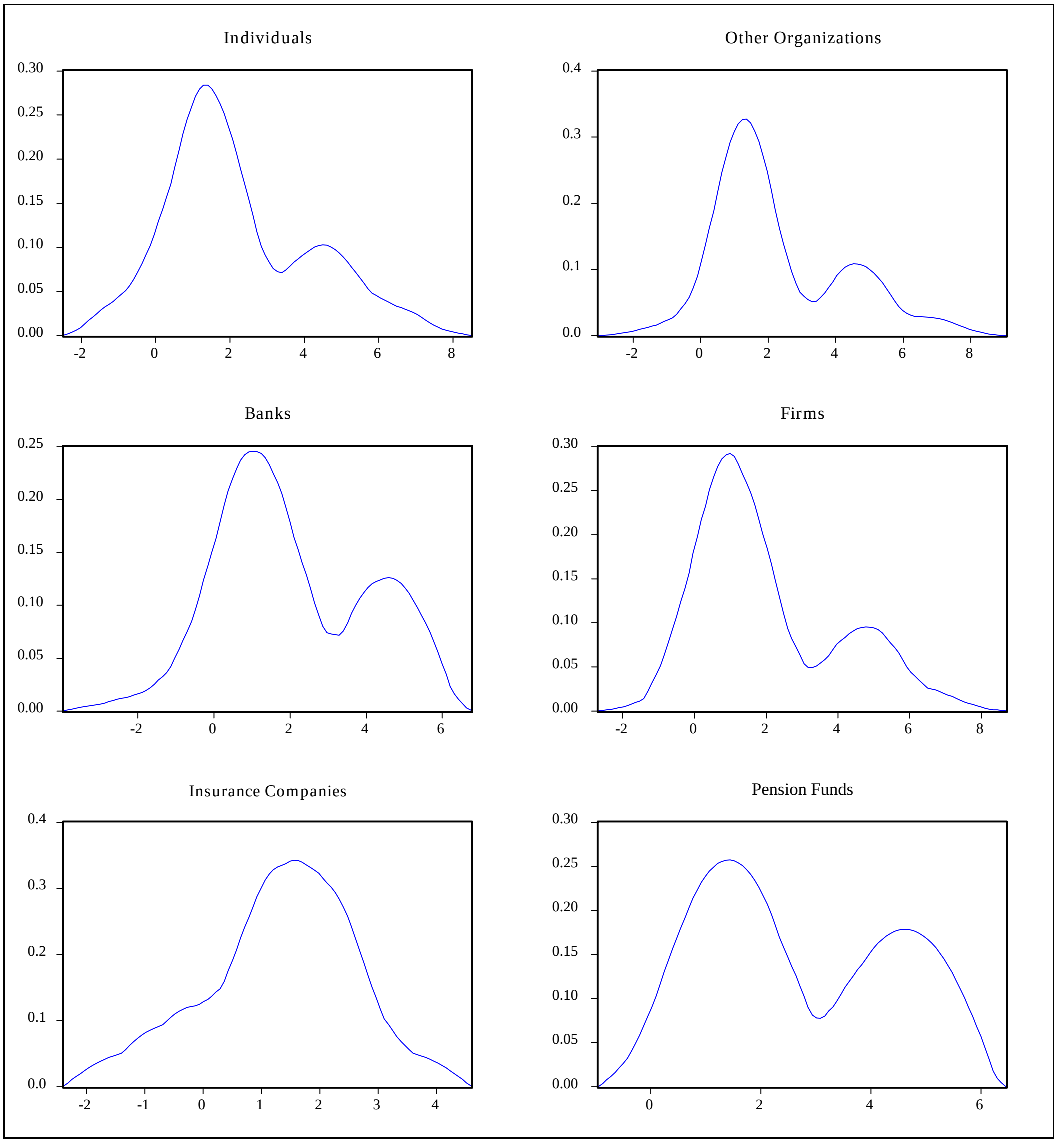
An important characteristic that a forecast error should have is that it should be not systematic (unpredictable). However, as figure 1 shows, and formal predictability tests would confirm, this is hardly the case. In fact, as we will show later, the underestimation bias is always present.

2.2 Are they all the same?

Given that in our database we can follow different individuals through time and through affiliation, we can evaluate whether or not there is any systematic difference among groups. This will unable us to verify whether they is a group of forecasters that dominates another.

[1] Despite being asymptotically valid, the tests of equalities (in mean and median) assume independence. More formal tests will be developed later.

Figure 2
Function of Density of and for Group



Note: The unconditional densities were estimated using the Epanechnikov kernel and the bandwidth selection proposed by Silverman (1986).

Figure 2 and Table 2 show the estimation of the unconditional density functions and equality tests for means among six groups. As can be observed all of the densities (with the sole exception of insurance companies) present evidence of bimodality.[2] Table 2 also shows a test for equality of the means among groups

[2] As Table 2 shows, there were only ten forecasts made by insurance companies in the whole sample. Thus, as will be shown later, this group does not refute the existence of bimodality in the other groups and on the aggregate.

(both for absolute and relative errors). As the null can not be rejected at standard levels of significance, for every practical purpose, there is no statistical difference among the groups. Thus these forecasters tend to commit the same type of errors both in terms of direction as well as magnitude.

Table 2
Descriptive Statistics of *e* and *r* by Group

Group	Observations	*e*		*r*	
		Mean	**Deviation**	**Mean**	**Deviation**
Individuals	353	2.183	0.101	1.469	0.030
Other Organizations	269	2.249	0.115	1.474	0.026
Banks	58	1.999	0.247	1.439	0.057
Firms	139	1.986	0.163	1.399	0.031
Ins. Companies	10	1.325	0.373	1.386	0.123
Pension Funds	28	2.675	0.310	1.514	0.047
Total	857	2.165	0.065	1.457	0.016
		Test	***P-value***	**Test**	***P-value***
ANOVA		1.252	0.283	0.668	0.648

Notes: Deviation = Standard error of the mean. ANOVA = Test of equality of means whose asymptotic distribution is of an F test with 5 degrees of freedom in the numerator and 851 in the denominator.

The tests conducted up to now do not take into consideration the temporal characteristics of the forecasts; that is, we considered the forecast errors unconditionally. Although it is clear that we should ask the forecast errors to be non-systematic, as already noticed, in practice all are. As was shown when analyzing the unconditional properties of the forecast errors, all of them tend to underestimate the growth rate of the economy. Thus, we will see if we can learn something more about this bias when analyzing the data conditionally.

A related question that can be asked is if there is any “leader” that others follow when conducting forecasts. In this case we understand as a “leader” to a group that precedes (statistically) another. Once we learned (from Table 2) that they is no statistical difference on the average forecast among groups, we computed the average forecast error of each group for each period and conducted a Granger causality tests. The results (available upon request) show marginal evidence of unidirectional causality from the forecast errors of the independent forecasters to other organizations.[3] This evidence (once again, marginal) suggests that the individual forecasters act as “leaders” and that the projections of the other organizations tend to “follow” them; thus, incurring in the same errors as the first ones.

2.3 When do they make fewer mistakes?

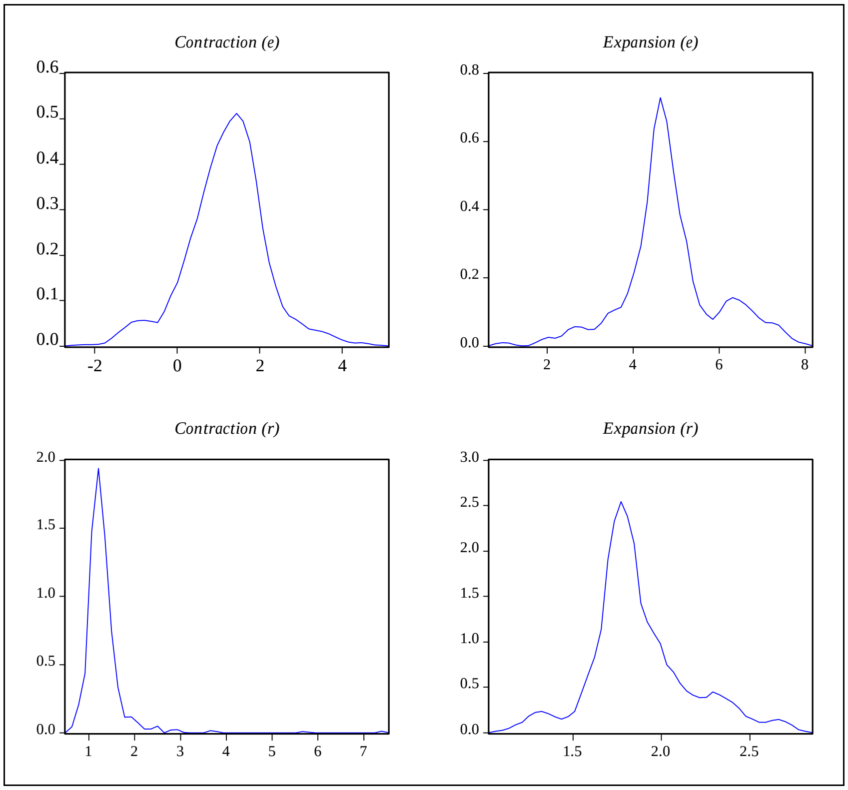
An undesirable property in the prediction errors is for them to be systematic, implying that the forecasters do not “learn” from their mistakes. Although we observed that there is an important underestimation of the growth rate of the economy, now we want to learn if there is pattern in this bias. One way to test whether or not there are systematic errors is to use conventional white noise tests such as the ones by Breusch and Pagan or Ljung and Box’s *Q* test. The results of these tests (not reported) strongly reject the null of white noise forecast errors when conducted to each forecaster, to a group of forecasters (average on a group), and to all of them (average of all forecasts for each period).

The fact that they make systematic mistakes does not help us to understand when and why they do that. We will immediately discuss the first issue (when) and later we will advance some reasons why.

As the evidence shows, the forecasters do not only make systematic mistakes, but the unconditional distribution of the forecast errors is bimodal. This fact can be easily explained when observing the forecast errors by year. If we associate a “contraction” with a year in which the GDP grew less than its average (7.7%) and an expansion with its complement, we observe that the forecast errors are, in addition to systematic, asymmetric. That is, the absolute (relative) errors despite being positive (exceeding 1) in all the phases of the cycle, they are smaller in the “contractions” than in the “expansions”.

[3] Most of the individual forecasters are academics in different of universities, while the forecasters in other organizations correspond generally to producer associations.

Figure 3
Density Functions of *e* and *r* by Phase of the Cycle



Note: The unconditional densities were estimated using the Epanechnikov kernel and the bandwidth selection proposed by Silverman (1986).

Table 3
Descriptive Statistics of *e* and *r* by Phase of the Cycle

Group	Observations	*e*		*r*	
		Mean	Deviation	Mean	Deviation
Contraction	629	1.195	0.038	1.307	0.018
Expansion	228	4.842	0.072	1.872	0.018
Total	857	2.165	0.065	1.458	0.016
		Test	*P-value*	Test	*P-value*
ANOVA		2266.175	0.000	323.296	0.000

Notes: Deviation = Standard error of the mean. ANOVA = Test of equality of means whose asymptotic distribution is of an F test with 1 degrees of freedom in the numerator and 855 in the denominator.

Figure 3 and Table 3 summarize this evidence. When we condition the estimation of the densities of the forecast errors to the "phase of the cycle" we now obtain unimodal distributions, although in both cases they are biased towards underestimation (as the simple mean tests suggest). Despite this, the forecast errors are asymmetric, in the sense that the underestimation is smaller in a “contraction” than in an “expansion”. Thus, on average, the forecasters underestimate the growth rate of the economy by more than one point during the “contractions” and by close to five points during the “expansions”.

Should this give us any comfort? Very little, because from the results reported it is easy to verify that that the variation coefficient of *e* is 2.3 times greater during a “contraction” than on an “expansion” (this coefficient is 3.6 times greater in the case of *r*). Summarizing; the forecasters are unnecessarily pessimistic in all the phases of the cycle, but particularly so during the “expansions.” The forecast errors (and therefore their projections) are more volatile (in relative terms) during a contraction that during an expansion.

2.4 Do they learn from they mistakes?

Even though the forecasts tend to present systematic biases towards underestimation, we could ask ourselves if the forecast errors tend to diminish when the forecasts are made closer to the period of projection. Given that *Estrategia* conducts several surveys during a given year, we construct a series that measures the distance (in months) between the period where the forecast was made and the period for which that forecast was made. Denoting the resulting variable by *L*, Table 4 shows the results of a regression between the forecast errors and *L*.

The regression displays several interesting features. Given that the forecast errors are systematic and asymmetric with respect to the phase of the cycle) we include a dummy variable that controls for this factor. As seen, this variable is highly significant for both, the absolute and relative errors. If the forecasters learned from their mistakes, we would expect the coefficient associated with *L* to be positive given that the further away from the realization of the series (the greater the value of *L*) we would expect the forecasters to be less accurate. It turns out that once heteroskedasticity and autocorrelation consistent (HAC) estimates of the variance of the parameters are used this variable is not statistically significant for standard levels of significance (the associated *p*-values for *e* and *r* are 0.13 and 0.79 respectively). The inclusion of HAC estimates for the variance covariance matrix is justified because the forecast errors display persistence and because White's heteroskedasticity tests (not reported) suggest its presence.[4] On the other hand, the coefficient associated with *L* for the regression on *r* is not significant and of the "wrong" sign.

Table 4
Linear Regression Models for *e* and *r*

	e			***r***		
	Parameter	**Deviation**	***p*-value**	**Parameter**	**Deviation**	***p*-value**
Constant	0.866	0.213	0.001	1.335	0.120	0.001
L	0.035	0.023	0.130	-0.003	0.011	0.787
D	3.645	0.172	0.000	0.565	0.048	0.001
	R^2=0.729 SER=0.985			R^2=0.273 EER=0.407		

Notes: Deviation = Standard Error of the parameter computed using the HAC variance-covariance matrix. *L* = Months between the projection and the realization of the series. *D* = Dummy variable that adopts the value of 1 in "expansion" and 0 in "contractions". R^2 = Adjusted R^2. SER = Standard Error of the Regression.

2.5 Why do they err that much?

Up to now we showed that the forecast errors are systematic and asymmetric. Here, we present a tentative explanation of this phenomena may occur. A reasonable search of variables that may explain the asymmetry found in the forecast errors is to see if there is any variable that has a different behavior according to the phase of the cycle. A natural candidate is of course the interest rate.

[4] As a matter of fact, the *p*-value associated with White's heteroskedasticity test for the regressions on *e* and *r* are 0.002 and 0.010 respectively. It is worth noticing that *L* turned out to be significant in "explaining" the squared residuals (with a positive sign). Estimations with Weighted Least Squares (using *L* as the weight) were also performed without changing the results of Table 4 significantly. This would imply, that even though they tend to make the same mistakes regardless of how near the effective realization of the variable is, at least their forecast errors tends to look alike due to the reduction in variance.

From an intertemporal perspective, the real interest rate is simply a relative price (between consumption today and tomorrow). Thus, if the economy is in a "contraction" that the agents perceive as transitory, their willingness the smooth their consumption stream would (generally) brings pressures for the interest rates to rise. Thus, it is not uncommon to find (weak) negative contemporary correlations between the growth rate of the economy and interest rates.

Even though the previous discussion applies to real interest rates, it is not uncommon for it to be translated to nominal or imperfectly indexed interest rates. In Chile, the Central Bank has a short-term instrument (an imperfectly indexed interest) called the PRBC that is usually taken to consider the stance of the monetary authority. In fact, this interest rate presents a mild negative contemporary correlation with the growth rate of GDP.

Table 5 shows the results from incorporating this variable in a regression for *e* and *r*. As can be observed, this variable is highly significant (at least in explaining the variation of the absolute errors) displaying a negative coefficient. This means that when this instrument increases, the forecast errors tend to diminish. This is congruent with the asymmetry found previously.

A misguided interpretation of these results would grant some type of economic causation from PRBC to the growth rate of the economy. This interpretation is not correct because the PRBC (as any other interest rate) is a good leading indicator of the expectations of the evolution of the economy; this mere fact does not bring any economic causation from one variable to the other. Thus, statistical precedence does not necessarily imply economic causation (see Chumacero, 1998a for a detailed discussion).

The more reasonable explanation of these results goes precisely the other way around. Recalling that the "dependent variable" is the forecast error (not the growth rate of the economy), and that we already controlled for the phase of the cycle, the results tend to show that the forecasters tend to grant an unjustified influence to the monetary authority's stance in having real effects.

Table 5
Linear Regression Models for *e* and *r*

	e			***r***		
	Parameter	**Deviation**	***p*-value**	**Parameter**	**Deviation**	***p*-value**
Constant	3.765	0.610	0.001	1.608	0.220	0.001
PRBC	-39.352	9.168	0.001	-4.606	3.577	0.198
D	3.423	0.157	0.001	0.539	0.043	0.001
	R^2=0.764 SER=0.920			R^2=0.281 EER=0.404		

Notes: Deviation = Standard Error of the parameter computed using the HAC variance-covariance matrix. *L* = Months between the projection and the realization of the series. *D* = Dummy variable that adopts the value of 1 in "expansion" and 0 in "contractions". R^2 = Adjusted R^2. SER = Standard Error of the Regression.

3 Do we need them to make forecasts?

Once we analyzed the statistical properties of the forecast errors, we ask ourselves if there is any way in which we can do better than they do. The metrics used to define what "better" means are the Mean Square Error (MSE) and the Root Mean Square Error (RMSE). Thus a "better" forecast will mean a lower MSE and a lower RMSE.

The exercise that is reported in this section corresponds to a comparison of the MSE and the RMSE of the forecasts reported in *Estrategia* and those of a "naive" forecaster. This forecaster (presumably in contrast with the others) will not have any type of "model" with which to elicit its forecasts. The characteristic of this forecaster will be that he always provides the forecast of constant growth rate for the whole period (1986-1997). We then compare the resulting MSE and RMSE of these forecasts with does of the other forecasters.

Table 6 shows the results of these comparisons. The column denoted by *e* reports the ratio of the MSE and RMSE of the forecasts of a given group and those of the "naive" forecasters. Thus, values that exceed 1 imply that the "naive" forecaster dominates a given group. For this exercise, the forecast of the “naive” forecaster corresponds to the average growth rate for the whole period (7.7%). As can be shown no forecaster will do better than the “naive” forecaster; even if we construct a “Best Forecaster” which does not correspond to any particular forecaster but to the observation that is the closest to the actual realization of the GDP, our “naive” forecaster does better. One may object this exercise on the grounds that our “naive” forecaster is not that naive after all because we “knows” what the average growth rate of the economy turned out to be. The column denoted by *Limits* shows the ranges of constant forecasts that for which our “naive” forecasters would do better than a given group. This means that, for example, any constant forecast on the range of [5.65,9.48] would render a lower MSE than the average of all the forecasts.

Table 6
Comparison of Projections

Group	***e***		***Limits***	
	A	**B**	***i***	***s***
Total	1.686	1.298	5.65	9.48
Individuals	1.763	1.328	5.55	9.58
Other Organizations	1.975	1.405	5.27	9.85
Best Forecaster	1.172	1.082	6.59	8.54
Worst Forecaster	2.629	1.622	4.60	10.53

Notes: A = Ratio between the MSE of a group and the MSE of the “naive” forecaster. B = Ratio between the RMSE of a group and the RMSE of the “naive” forecaster. Limits = Intervals of constant forecasts for the whole period that dominate in MSE and RMSE to each group. *i* = Lower bound of the interval. *s* = Upper bound of the interval. Total, Individuals, Other Organizations = Comparison with the average of forecast error for each periods committed by a group. Best Forecaster = Comparison with the forecast error that is closest to the actual value of the growth rate of GDP for each period. Worst Forecasters = Comparison with the forecast error that is farther to the actual realization of the series.

Obviously, the fact that a “naive” forecaster has such a wide range for beating all the forecasters in the sample is not only a strong but uncomfortable result. Chumacero (1998b) develops a simple time series model that is indeed able to provide forecasts errors than dominate our “naive” forecaster; thus, by extension, performing better than any of the forecasters included in the sample.

4 Should we care?

As shown, this paper presents statistical evidence that shows that the forecasters systematically underestimate the true growth rate of the Chilean economy for the period comprehended between 1986 and 1997. Furthermore, the magnitude of this bias tends to be correlated not only with the phase of the business cycle but also with the monetary authority's stance.

Said that, is there any practical reason why we should worry about the poor performance of the forecasters? If their forecasts are not taken into consideration for "real-life" decision-making, the answer should clearly be negative. Even if the private sector abstracted from this noise there is still a reason why we should worry about this bias towards underestimation and it has to do with the public sector. Even though they were not considered in the sample, both the Central Bank of Chile and the Finance Ministry provide "forecasts" of the performance of the economy once a year. These forecasts tend to have the same bias of the forecasts analyzed in this document. In the past years, the authorities have always underestimated the growth rate of the economy by margins that are similar to those of the forecasters included in our sample.

A simple corroboration of this fact is that the public sector has been not only been running a consistent pattern of fiscal surpluses, but also that the observed revenues by the end of a fiscal year are systematically superior to the revenues that were forecasted in the budget at the beginning of the year.

This is one potential avenue through which these forecasts could matter. Consider for example that the authority has a preset sequence of expenditure that it wants to finance. The marginal tax rates (say on income) would be set accordingly. If the authority underestimates the growth rate of economy it would set rates that would be higher than those necessary for financing that sequence of expenditures. Once the true (higher) growth rate occurs, the authority finds itself with an "unexpected" surplus because, given that tax rate it collects more than what was originally planed. Of course, the welfare costs of such an underestimation are clear: lower tax rates will be sufficient to finance that sequence of expenditure making the taxpayers happier in the process.

References

Chumacero, R. (1998a). "A Toolkit for Analyzing Alternative Policies in the Chilean Economy." *Manuscript*. Universidad de Chile.

Chumacero, R. (1998b). "Qué Dicen los Expertos?" *Manuscript*. University of Chile.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.