FORECASTING ACTIVITY USING SENTIMENT INDICATORS:

The Case of Chile

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN ANÁLISIS ECONÓMICO

Alumno: Matías Nicolás Rojas Alabarce
Profesor Guía: Pablo Matías Pincheira Brown

Santiago, Marzo 2016

Abstract

In this article we evaluate, in several dimensions, the ability of two sentiment indicators, consumers and business, to forecast year-on-year variation of Chilean activity. We do in-sample and out-of-sample exercises to evaluate predictive capacity. In out-of-sample exercises, when we predict activity using a constant we find that the business confidence indicator (BCI) have the capacity to improve forecasts, the economic perception index (EPI) do not shown predictive capacity in a naïve context. Adding a univariate structure, the results continue to show predictive ability for the BCI, the variable improve forecast for horizons 1 to 12 month ahead, for the EPI the results show predictive ability in medium term forecast, 9 and 12 months ahead. When we use the model proposed by Urrutia and Sanchez (2008) (USM), BCI continues to show predictive ability, but by itself, on average, does not deliver better forecasts that USM. For EPI we find predictive ability in out-of-sample exercises for medium term forecast. The hit rate exercises shows that BCI and EPI correctly predict changes in direction of activity in most horizons. We conclude that business confidence indicator can be use as a leading indicator of Chilean activity. A contribution of this paper is use Clark and West (2007) test in iterative method of forecast.

Index

I. Introduction

Forecast activity is relevant to both private and public sector. For Central Banks is important to have accurate predictions about the future path of economic activity, both at long and short horizons, in order to improve the analysis required for monetary policy decisions. This is especially important in those countries which have the purpose of price stability and inflation targeting, since the growth of activity has been traditionally reported as having an impact on inflation. Also, in assessing fiscal sustainability it is crucial to have a good forecast of the future path of national output. From the private perspective can also be relevant as long as be able to orient decisions of national or sectorial investment.

Notwithstanding the above, activity forecasts are, unfortunately, no as accurate as one would wish, especially during times of turmoil (Chauvet and Potter, 2013). So the search for new methods or variables to increase forecast accuracy is needed.

In this article we evaluate, in several dimensions, the ability of two monthly sentiment indicators to forecast activity in Chile. We place special attention to the role these sentiment indicators may be playing to improve the accuracy of activity backcasts and nowcasts. This is important because is well known that official activity data is released with some lags in several countries, leaving economic authorities in blind when making policy decisions in real time. Our analysis focuses on models to forecast year-on-year activity in Chile[1]. The main idea is to prove that these sentiment indicators can be used as leading indicators of Chilean activity.

We find that Business Confidence Indicator (BCI) have predictive ability beyond contained in good univariate and multivariate benchmarks. The BCI improve forecast for horizons 1 to 12 month ahead, especially truth in short term forecast, backcast and nowcast. The results for Economic Perception Index show predictive ability in medium term, horizons 9 and 12 month ahead. The hit rate results show that this two sentiment indicator have ability to forecast change in direction of activity growth.

[1] Hereafter when we talk about activity we refer of year-on-year variation of IMACEC.

A contribution of this paper is use Clark and West (2007) test in iterative forecast, the asymptotic theory of the test works with direct forecasts. To our knowledge this is the first attempt to use this test for different horizons in iterative forecasts.

The rest of the paper is structured as follows. A brief literature review and economic theory are presented in section II. In section III we describe the data, section IV presents the econometric approach to evaluate predictive ability. In section V we show the models used to build our forecasts. The results of our forecasting exercises are found in section VI and finally section VII concludes.

II. Literature review and economic theory

Literature review

In this section we review the literature related to 3 topics: Forecasting activity, sentiment and leading indicators. It is necessary to review the literature about forecasting activity in order to build a good benchmark for our model.

Some definitions first: What is a sentiment indicator? It is a numerical indicator designed to show how a group (e.g. investors or consumers) feels about the market, business environment or other factor. What is a leading indicator? It is a measurable economic factor that changes before the economy starts to follow a particular pattern or trend (Hüfner and Schröder, 2002). Leading indicators are used to predict changes in the economy, but they are not always accurate.

The literature about forecasting activity is extensive. The variables that have proven ability to forecast activity are interest rates and spreads (yield curve), stock prices, monetary aggregates, expectations surveys and indexes of leading indicators (Stock and Watson, 2003; Huang et al, 2006; Stock and Watson, 2002). About the methods, literature finds that DSGE are comparable or slightly above of the VAR and BVAR, but not significantly better than the simple univariate benchmark (Chauvet and Potter, 2013).

In Chile, this literature is quite new and under-researched. Urrutia and Sanchez (2008) use data from generation of electricity to build 1 month ahead forecasts for IMACEC (strictly speaking a backcast), they find that their model beats SARIMA models in out-of-sample exercises. Calvo and Ricaurte (2012) build two composite coincident indexes, based on works of others authors (Bravo and Franken, 2001; Firinguetti and Rubio, 2003; Pedersen, 2008), with financial data and OECD methodology, and they use Urrutia and Sanchez model (USM, hereafter) as benchmark. The authors find that composite coincident index models have problems in presence of exogenous events, e.g. an earthquake, due to its inertia, this does not occur with USM, which adjusts faster. In individual terms, composite index models cannot beat the model of generation of electricity. Combination of these two composite index and USM displays the lowest root mean squared prediction error.

The literature about sentiment indicators is extensive. Several authors have found that sentiment indicators have power to forecast recessions, stock prices, consumer spending, activity and industrial production (Christiansen et al., 2013; Hengelbrock et al., 2011; Ludvigson, 2004; Posta and Pikhart, 2012; Hüfner and Schröder, 2002; among others). Christiansen et al. (2013) find that Purchasing Managers Index (PMI), developed by the Institute of Supply Management, is by far the best single recession predictor in both in-sample and out-of-sample analyses comparing with others sentiment indicators and financial variables. Posta and Pikhart (2012) study the use of sentiment indicators to forecast GDP for EU economics. They find out that the relationship between ESI (Economic Sentiment Indicator, published by the European Commission) and GDP may be exploited in relatively stable times while the relationship may be quite distorted when an economy is hit by unexpected shocks.

In Chile, Pincheira (2014) use business confidence indicators (BCI) to predict aggregate and sectoral employment. He finds forecast power in aggregate employment and construction sector. Echavarría and González (2011) use the business confidence indicator and other variables (real exchange rates, exports, imports, stock index, etc.) in a model of dynamic factors to forecast activity in Chile (IMACEC), the results show that dynamic factor model in some cases is better than a univariate model and expectations survey.

Finally, the literature about leading indicators for Chile is related to build composite coincident indexes, but leaving explicit out-of-sample forecasting exercises aside. For example, Bravo and Franken (2001), Firinguetti and Rubio (2003) and Pedersen (2008) build composite coincident indexes using financial variables that leads activity, these authors do not make an out-of-sample exercise to evaluate forecast power, but, as we mentioned, Calvo and Ricaurte (2012) use these indexes to only find that they cannot beat USM.

Given this review, we conclude that a good benchmark (and easy to build) is the model developed by Urrutia and Sanchez (2008), we extend this model to forecast activity for more than one step ahead.

Economic Theory

The economic theory related to this article is survey expectations and the rational expectations hypothesis (Pesaran and Weale, 2006). Multiple-equilibria macroeconomic models suggest that consumers' and investors' sentiment (sunspots and animal spirits, respectively) about the state of the economy may be an important independent factors for business cycles. The idea of self-fulfilling prophecy, i.e. the expectations that individuals have on the economy can influence their behavior, come from there. For example, if people expect more inflation, they may accelerate their purchasing to the degree that it has an inflationary effect.

In this paper, we use consumers and business sentiment indicators; the theoretical question that arises is what are the transmission channels, i.e. how the feeling of individuals will affect the path of the product. In general, it argues that the dominant channel is the self-fulfilling prophecy, but there may also be some additional information held by investors that affect the fundamentals of the economy, such new investment plans have not yet become public. Chauvet and Gou (2003) study the interrelations between waves of optimism and pessimism and subsequent economic fluctuations in US, finding that waves of pessimism may have played a nontrivial role for the 1969-70, the 1973-75, and the 1981-82 recessions.

While the theoretical framework is always important, the purpose of this paper is not to find the transmission channels but rather test predictive accuracy of sentiment indicators that are produced in Chile, in order to show that these indices can be used as leading indicators. Future research could test how the movements of the indicators, not based on fundamentals (sunspot), affect the Chilean economy.

III. Data

In this section we describe the data. For the analysis we have monthly frequency data for the period January 2004-February 2015 (134 observations). Our measure of activity is the monthly index of economic activity (IMACEC) revised and published by the Central Bank of Chile, the IMACEC is a representative index of economic activity in Chile, covering about 80% of the goods and services that compose the country's GDP and emulating therefore part of his behavior; this allows to have information on economic activity more frequently than the official data. Is published with a lag of 35 days[2].

For this paper we use two sentiment indicators, the monthly index of business confidence (BCI) developed by ICARE-UAI[3], the index is based on a monthly survey where they interview about 600 business executives from manufacturing, mining, construction and trade sectors. The sample consists of a panel of companies, so that the same units are consulted at every opportunity. The business executives are asked questions about activity, future activity, demand, stock, sellings, inflation, among other questions. The index is built up from responses balances, using the same methodology that confidence indicator from the Institute for Supply Management (better known as PMI)[4]. The main indicator (BCI) is constructed as a weighted average of four sectoral confidence indicators (manufacturing, mining, construction and trade). The weights of the BCI correspond to the share of these sectors in GDP. Index ranges from 0-100, with a natural barrier of 50, this means that a diffusion index greater than 50 reveals an "optimistic" or "favorable" confidence level with respect to the variable analyzed and, on the other hand, reveals an "unfavorable" confidence, if it falls into the lower range of the "neutral barrier". The index is published at the beginning of the following month (5 days after the last day of the month).

Our second sentiment indicator is the monthly index of economic perception (or economic perception index, EPI) developed by Adimark-GFK[5], it is a composite index, which is calculated from the combination of public responses to 5 questions[6]. A sample of 1000 men

---

[2] All data in this study were obtained from the database of the Central Bank of Chile.
[3] We use BCI, business confident indicator, in Chile this index is called IMCE for its acronym in spanish.
[4] For more details go: http://www.icare.cl/imce-2014/ficha-tecnica
[5] We use EPI, economic perception index, in Chile this index is called IPEC for its acronym in spanish.
[6] For more details go: http://www.adimark.cl/es/estudios/documentos/ipec%20adimark-gfk%20ene15.pdf

and women aged 18 years and more living in the main cities of Chile is surveyed. These questions attempt to measure the perception of the public about current personal financial conditions, current national economic conditions, future national economic conditions in the short term (one year), future national economic conditions in the long term (5 year) and expectations consumption of household items. Same as with the BCI, this index ranges from 0-100, with a natural barrier of 50, this means that a diffusion index greater than 50 reveal a confidence level "optimistic" or "favorable" with respect to the variable analyzed and, on the other hand, reveals an "unfavorable" confidence, if it falls into the lower range of the "neutral barrier". The index is published at the beginning of the following month (5 days after the last day of the month).

We include a variable related to generation of electricity to build our benchmark based on the work of Urrutia and Sanchez (2008). As the authors mention in their paper, they assume that the dispatches of electrical energy are equivalent to the total electricity consumption in the country. The unit of measure used is gigawatt / hours (GWh) and the source of information is the Center for Economic Load Dispatch (CDEC, for its acronym in Spanish). The data is published at the beginning of the following month.

Table 1 shows the main descriptive statistics of the aforementioned variables, activity and Energy are displayed in annual variation, and BCI is at levels and the same goes for EPI.

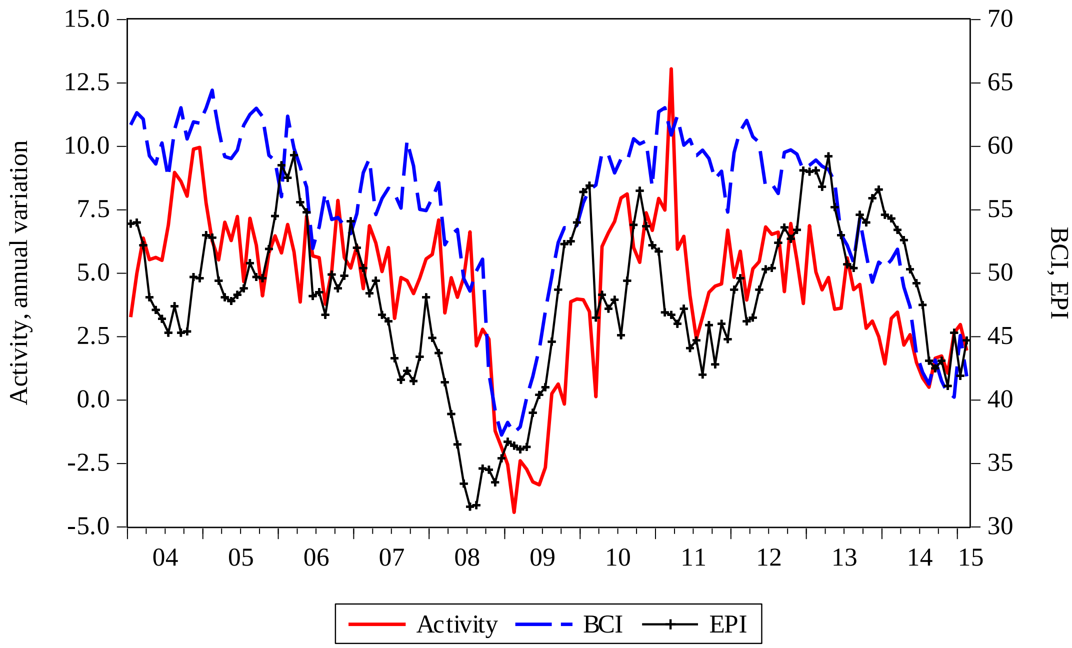
Table 1. Descriptive Statistics

Statistics\Variable	Activity	BCI	EPI	Energy
Mean	4.48	54.97	47.99	4.11
Median	4.84	56.93	48.45	4.26
Maximum	13.04	64.42	59.30	16.55
Minimum	-4.43	37.23	31.60	-7.80
Std. Dev.	2.88	6.74	6.24	3.20
Observations	134	134	134	134

Source: Author’s Elaboration

Figure 1 helps to motivate the predictive exercises of the upcoming sections. In this graph we see that the BCI has a significant ability to anticipate the recovery after the Great recession of 2009. The EPI seems to anticipate the fall previous to the great recession. Also, Figure 1a (see appendix) shows activity and energy, we can see that both series have similar behavior.

Figure 1: Activity, Business Confidence Indicator and Economic Perception Index, January 2004-February 2015



Source: Author's elaboration, Central Bank of Chile, ICARE-UAI, Adimark-GFK.

Figure 2 helps to motivate our analysis as well. In this chart we see the cross correlogram between activity and BCI, and activity and EPI. The correlation structure makes it evident that indexes anticipate in several months variations in activity. This is particulary true in BCI but less clear in EPI.

The existence of correlations do not give a measure of forecast error when we predict activity with BCI or EPI. Even more, cross correlogram does not allow us to evaluate predictive capacity of BCI or EPI in relation with others natural predictors of activity (energy for example).

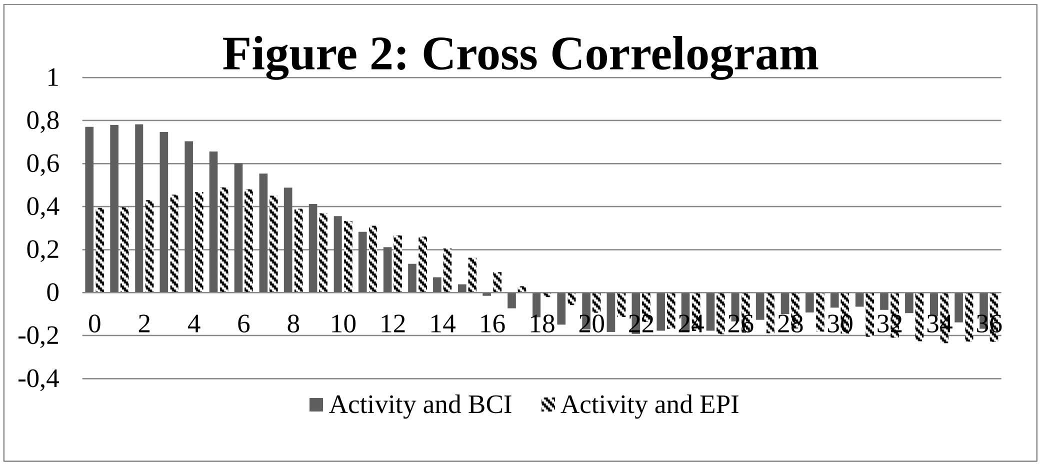
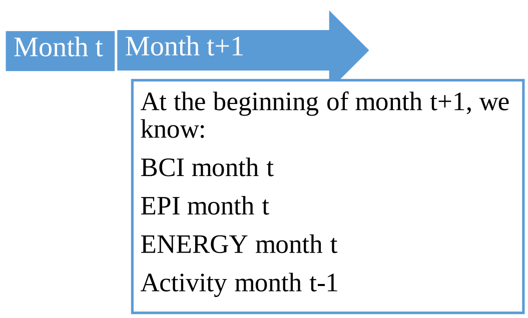


Figure 2: Cross Correlogram

Source: Author’s elaboration.

Figure 3 shows the publication schedule, as we mentioned, our variables have different release dates. This is important to the aim of this paper, as we want to test whether confidence indicator can nowcast, backcast and forecast activity.

Figure 3: Publication Schedule



To the out-of-sample exercises our evaluation scenario will be the monetary policy meeting, for Chile takes place in the second week of the month. Then the president and board members of the Central Bank, analysts and investors know BCI, EPI and Energy of the month t-1 and activity of the month t-2. So our first forecast of activity using BCI or EPI is going to be a “backcast” (activity in month t-1), our second forecast of activity is going to be a “nowcast” (activity in month t), and the followings are going to be forecasts.

IV. Predictive Evaluation Strategy

Our predictive evaluation strategy consists of exercises both in-sample and pseudo out-of-sample. We describe these exercises next:

1. In-Sample exercises

First, we perform exercises based on an ADL model (autoregressive distributed lag model) which seeks to determine if the BCI or EPI has the ability, in the absence of other variables or lags, to forecast annual variation in activity. Second, we make a bivariate Granger causality test between activity and BCI, and activity and EPI. We build a VAR of order 12 between activity and BCI or EPI, and then we carry out a Granger causality test to evaluate if BCI has predictive power, in addition to that contained in the lags of activity variable by itself.

Third, we consider the univariate family of SARIMA processes to find a good representation of activity. Using this good univariate representation, we explore the ability of BCI or EPI to improve in-sample fit to augment these specifications so that variable appears explicitly in the expression. Then using either a t or F test we evaluate the statistical significance and additional predictive capacity that our confidence indicators add to the univariate SARIMA representation. We also make these 3 exercises with energy data as an exogenous predictor.

2. Out-of-sample exercises[7]

In our out-of-sample exercises first we consider a naïve model of activity, by predicting activity based only on a constant as benchmark and we augment this model with BCI or EPI. Then we consider a univariate model for activity as benchmark and its extended counterpart with corresponding BCI or EPI. Finally, motivated by the work of Urritia and Sanchez (2008) we build a third benchmark based on a univariate model of activity augmented by energy data. In this point, we make a difference with the work of them because we do not use seasonally adjusted data in our analysis, instead we model the seasonal pattern with seasonal moving average terms.

To explain the out-of-sample exercises, suppose we have a total of T+1 observations of activity (Y_t), with that we generate a sequence of forecast h-step ahead (F(h)) estimating

[7] In this article, we use revised data. Therefore, a natural extension for further research is to evaluate the predictive content of our sentiment indicators using a data available in real time.

models in recursive windows of variable size (also we use a fixed rolling window, the results do not change significantly). Consider that the first window of estimation has a generic size of R. To make clear, to generate the first forecast h step ahead we estimate a model with the first R observations of the sample. Then we build the first forecast with information available until the observation R and we compare that with the realization Y_{t+h}. In a second stage we estimate models with the second recursive window which consider the first R+1 observations of the total sample. We built new forecast h step ahead and compare with realization Y_{t+h+1}. We continued iterating to consider the last recursive window that contains the first T + 1-h observations. The forecast build with these estimators are compared with Y_{t+1}. Finally, we build a total of F(h) forecast h-step ahead, with F(h) satisfying R+(F(h)-1) +h=T+1. This way, we have:

$$F(h) = T + 2 - h - R$$

In this paper, we use a window size of 50 observations (R=50), then the first window covers from January 2004 to February 2008. This implied that we build a total of 84 forecast one month ahead covering the period March 2008-February 2015[8]. We use as a measure of predictive accuracy the Root Mean Squared Prediction Error (RMSPE) and the Mean Absolute Prediction Error (MAPE). Because both are population moments, we report their sample counterparts calculated as follows:

$$RMSPE = \sqrt{\frac{1}{F(h)} \sum_{t=R}^{T+1+h} \left(Y_{t+h} - \hat{Y}_{t+h|t}\right)^2}$$

$$MAPE = \frac{1}{F(h)} \sum_{t=50}^{T+1+h} \left|Y_{t+h} - \hat{Y}_{t+h|t}\right|$$

The variable $\hat{Y}_{t+h|t}$ represent the forecast of Y_{t+h} made with information available at time "t". For the out-of-sample exercises we show the RMSPE and MAPE ratios, this is the RMSPE (MAPE) of the augmented model divided by the RMSPE (MAPE) of the benchmark

[8] The number of forecasts declines as the prediction horizon increases. So, we only have 83 two months ahead forecasts covering the period April 2008-February 2015, 82 three months ahead forecasts covering the period May 2008-February 2015, so on.

model, if the ratio is lower than one means that the augmented model is more accurate than the benchmark.

To evaluate whether the differences in predictive accuracy are statistically significant, we proceed to compare our benchmark models with their augmented versions with either BCI or EPI (the naïve, univariate and augmented by energy benchmarks). We rely primarily on two different paradigms to base our statistical inference. The first of them it is attributed to Diebold and Mariano (1995) and West (1996). This strategy, and its t-type statistic, is referred as to "DMW test", hereinafter. The second paradigm is the approach recently proposed by Clark and West (2007), CW hereinafter.

According to the paradigm of DMW we evaluate the following null hypothesis:

$$H_0: E\left(\hat{d}_t(h)\right) \leq 0$$

Against the following alternative hypothesis:

$$H_A: E\left(\hat{d}_t(h)\right) > 0$$

In which:

$$\hat{d}_t = \left(Y_{t+h} - \hat{Y}_{1,t+h|t}\right)^2 - \left(Y_{t+h} - \hat{Y}_{2,t+h|t}\right)^2$$

And $\widehat{Y}_{1,t+h|t}, \hat{Y}_{2,t+h|t}$ denote h-step ahead forecasts generated by the two models under consideration[9]. Model 1 is the parsimonious or "small" model, while model 2 is the "big" model that nests model 1. In other words, if we restrict some parameters to zero, model 2 would be exactly equal to model 1.

We focused on one-sided tests because we are interested in detecting predictive superiority. Our null hypothesis assumes that the forecast generated by the nested models are at least as accurate as those generated by the larger model. On the contrary, our alternative hypothesis suggests that the large model forecasts are more accurate than the nested model forecasts.

In second place we focus in the paradigm of Clark and West (2007). The objective of this paradigm is evaluate nested models using out-of-sample forecasts. This objective is different

[9] The DMW test that we describe here can easily extend to others lost functions, such as the absolute value.

to DMW test, which consists of comparing predictive accuracy. CW test seeks to evaluate whether a set of variables are statistically significant, to this end it uses out of sample forecasts. In this context, we use CW test to evaluate statistically significance of a set of BCI or EPI variables which integrate as additional elements in a univariate specifications of activity.

CW test is usually interpreted in two different ways. First, it can be consider as an "encompassing" test. This means that the test evaluates whether it is possible to improve the predictive accuracy of the two models under evaluation by taking a weighted average of them. Other way to see CW test is to understand it as a test which allows comparing the predictive behavior of two nested models, through an "adjusted" comparison of the respective mean squared prediction errors (MSPE). The adjustment is made in order to introduce justice to these comparisons. Intuitively the test eliminates a term that introduces noise when a vector parameter is estimated which is equal to zero under the null hypothesis of equal MSPE. Clark and West test is built with the following core statistic:

$$\hat{z}_{t+h} = \left(\hat{e}_{1,t+h}\right)^2 - \left[\left(\hat{e}_{2,t+h}\right)^2 - \left(\hat{Y}_{1,t+h|t} - \hat{Y}_{2,t+h|t}\right)^2\right]$$

In which $\hat{e}_{1,t+h} = Y_{t+h} - \hat{Y}_{1,t+h|t}$ and $\hat{e}_{2,t+h} = Y_{t+h} - \hat{Y}_{2,t+h|t}$ represent the corresponding forecast errors. With some algebra is possible to show that the test of CW can be expressed as:

$$Adjusted - MSPE = \frac{2}{F(h)} \sum_{t=R}^{T+1-h} \hat{e}_{1,t+h}(\hat{e}_{1,t+h} - \hat{e}_{2,t+h})$$

This statistic is used to evaluate the following null hypothesis:

$$H_0 : E(Adjusted - MSPE) = 0$$

Against the following alternative hypothesis:

$$H_A : E(Adjusted - MSPE) > 0$$

Clark and West (2007) propose testing this null hypothesis with a one-sided t-test using asymptotically normal critical values. The CW test is used in direct forecasts, here we use iterative forecasts, for this particular case we do not know whether normal critical values will render a test with correct size. To deal with this problem we make Monte Carlo simulations of the out-of-sample exercises, to obtain the critical values that allow us to reject the null hypothesis at a significance level of 10%, 5% or 1% (see Appendix). To our knowledge the CW test has never been used for iterative forecasts, and then this is the first attempt to use this test for different horizons in iterative forecasts.

It is important to emphasize that both test, DMW and CW, are quite different. They are designed to test different null hypotheses which seek different objectives. DMW test is used in this paper in order to evaluate predictive accuracy of two different strategies. In the other hand, CW test seeks to evaluate if a model is more adequate than other in a context of nested models. As a consequence, it can be expected that the two tests deliver different results.

To complete our analysis and differing from traditional RMSPE comparison, a policymaker or researcher may also be interested in the ability of different forecasting models may have to correctly predict if growth are going up or down. We evaluate this dimension of our forecasting models by computing the hit rate i.e. the rate of correctly forecasting the direction of change in growth rate (See appendix for further explanation).

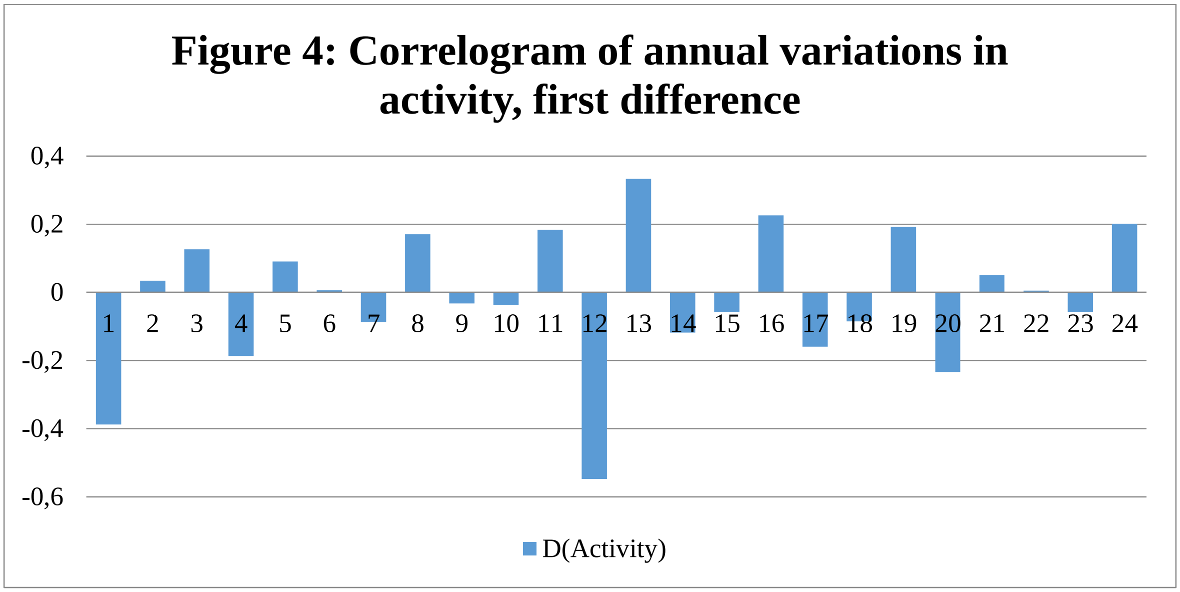
V. Forecasting Models

In this section we present with more detail the models that we use for our predictive analysis. Especially, we present ADL specifications, bivariate VAR and univariate SARIMA aforementioned. For the selection of models, we rely on the identification methodology presented by Box and Jenkins (1970). In some cases, we select parsimonious models despite their information criteria are greater than more complex models, because the windows estimation are not so large as to estimate more robust parameters, and in the out-of-sample exercises benchmark results are better with more parsimonious models, to make a fair comparison.

Is very intuitive to think that activity in levels has a seasonal pattern, some times of the year have a greater economic activity, end of year for example. But is less intuitive than annual variations still present a seasonal pattern. If we consider the logarithm approximation of the annual variation of the activity:

$$Y_t = 100[Ln(output_t) - Ln(output_{t-12})]$$

It is clear that seasonal additive terms disappear of the logarithm of activity by taking the difference in twelve months. However, seasonal multiplicative terms or with nonlinear specifications do not have to go away. Figure 4 shows the correlogram of the first differences of the annual variation in activity. We consider the first differences to eliminate low-frequency trend components of the sample. We see a large-scale autocorrelation in the first lag, then the magnitude decays in the following lags to return to a striking extent in the lag 12 and 13. This behavior is consistent with a seasonal pattern. For this reason, we consider an ADL model and a bivariate VAR both of order twelve as first analytical tools, waiting for the incorporation of the 12 lags allow to capture the seasonal behavior.



Source: Author's elaboration.

The specification that we used for ADL model is the following:

$$y_t = c + d_0 I_t + d_1 I_{t-1} + \ldots + d_{12} I_{t-12} + \varepsilon_t$$

Where y_t represent the annual variation of activity and I_t represent the index BCI or EPI. ε_t shock is a white noise. This first specification is used for evaluated statistically if BCI have capacity to predict annual variation of activity. The statistical significance of F-statistic gives us the answer to this question. Second, we carry out a Granger causality test based on the following bivariate VAR of order 12:

$$y_t = c + a_1 y_{t-1} + \ldots + a_{12} y_{t-12} + \ b_1 I_{t-1} + \ldots + b_{12} I_{t-12} + u_t$$

$$I_t = \alpha + \ f_1 I_{t-1} + \ldots + f_{12} I_{t-12} + g_1 y_{t-1} + \ldots + g_{12} y_{t-12} + w_t$$

As usual, shocks u_t and w_t correspond to a vectorial white noise with a positive definite variance-covariance matrix. Like we mentioned before I_t can be either BCI or EPI.

An alternative strategy to popular VAR modeling is using univariate models SARIMA (seasonal ARIMA), which as its name implies, explicitly modeling the seasonal components of the series in question. The choice of each SARIMA model for the series of activity, BCI and EPI is based on the criteria of identification of models proposed by Box and Jenkins (1970). This strategy involves analyzing autocorrelograms of the series to identify the orders of integration, the existence of seasonality and the maximum orders of the autoregressive and moving averages polynomial. Then in a second phase several SARIMA models that are consistent with the structure of the correlogram are estimate, finally we choose the model with all term statistical significance and residuals behave like a white noise, the final model is that with the lowest Akaike within the possible options. Table 2 below shows the modeling for each series.

Table 2. In-sample SARIMA and Out-of-sample specifications for Activity and BCI, EPI and Energy:

Activity	$y_t = c + \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \Theta_1 \varepsilon_{t-12} + \theta_1 \Theta_1 \varepsilon_{t-13}$
Out-of-sample specifications:	
Activity naïve benchmark	$y_t - y_{t-1} = c + \varepsilon_t$
Activity univariate ARIMA	$y_t - y_{t-1} = \beta(y_{t-1} - y_{t-2}) + \varepsilon_t + \vartheta \varepsilon_{t-12}$
Activity univariate ARIMA augmented by Energy	$y_t - y_{t-1} = \gamma(X_t - X_{t-1}) + \delta(X_{t-1} - X_{t-2}) + \varepsilon_t$ $\varepsilon_t = \rho \varepsilon_{t-1} + v_t - \alpha v_{t-12}$
BCI	$I_t - I_{t-1} = \varphi(I_{t-1} - I_{t-2}) + \epsilon_t + \theta \epsilon_{t-12}$
EPI	$I_t - I_{t-1} = \eta(I_{t-1} - I_{t-2}) + \epsilon_t + \omega \epsilon_{t-12}$
Energy	$X_t - X_{t-1} = \pi(X_{t-1} - X_{t-2}) + u_t + u_{t-12}$

Source: Author’s elaboration.

Table 2 shows specifications for in-sample analysis, SARIMA model, and for out-of-sample exercises. As mention before, in some cases we choose the parsimonious model rather than the model with the lowest Akaike, we do that because in out-of-sample exercises the result of parsimonious model outperforms more complex models. y_t correspond to the year-on-year variation of IMACEC, X_t is the year-on-year variation of energy data. I_t is BCI or EPI at levels. Notice that in the out-of-sample specifications we do not use a SARIMA models, the multiplicative lag, this is because the windows size and data are not big enough to allow us to have a good estimation of the multiplicative lag, and when we use ARIMA models the results are better off, i.e. a lowest MSPE for the benchmark models.

To build multiple horizons forecasts we use the iterative method that requires, for more than one-step-ahead forecast, the definition of an auxiliary model to generate forecasts of exogenous variable (BCI, EPI or Energy). That is the reason why in Table 2 also shows the univariate specifications we use to generate these forecasts. Marcellino, Stock and Watson (2006) prove that the iterative method outperform direct method to generate forecasts of output growth, this is the reason why we use iterative method in this paper.

To complete the econometric framework, our out-of-sample analysis focuses on compare a univariate model of activity with the same univariate model but augmented by BCI or EPI and their lags. As mention before, we have 3 benchmark, a constant model, a univariate ARIMA and a univariate ARIMA augmented by energy data (see table 3), inspired by the work of Urrutia and Sanchez. So we also compare the univariate ARIMA of activity augmented by energy with the same model augmented by BCI or EPI and their lags. Urrutia and Sanchez (2008) prove that a SARIMA model of activity augmented by energy performs better than a univariate model, so we want to test if BCI or EPI improves forecast analysis to include these variables on the model aforementioned. For example, we have the following model of activity, the univariate ARIMA benchmark:

$$y_t - y_{t-1} = \beta(y_{t-1} - y_{t-2}) + \varepsilon_t + \vartheta\varepsilon_{t-12}$$

Then we augmented this model to form this expression:

$$y_t - y_{t-1} = \beta(y_{t-1} - y_{t-2}) + \varepsilon_t + \vartheta\varepsilon_{t-12} + \beta_0 I_t + \beta_1 I_{t-1} + .. + \beta_j I_{t-j}$$

Notice that I_t can be BCI or EPI. Then we evaluate the joint significance of β coefficients both in-sample and out-of-sample. We include a contemporaneous variable I plus j lags. We can include the contemporaneous variable because of the publication schedule and the evaluation scenario explained before. Table 3 shows the augmented models for different benchmarks:

Table 3. Out-of-sample specifications:

Activity naïve benchmark augmented by …	
BCI	$y_t - y_{t-1} = c + \phi(I_t - I_{t-4}) + \varepsilon_t$
EPI	$y_t - y_{t-1} = c + \nu(I_t - I_{t-5}) + \varepsilon_t$
Activity univariate ARIMA benchmark augmented by …	
BCI	$y_t - y_{t-1} = \beta(y_{t-1} - y_{t-2}) + \lambda\left(\frac{I_t - I_{t-5}}{I_{t-5}}\right) + \varepsilon_t + \vartheta\varepsilon_{t-12}$
EPI	$y_t - y_{t-1} = \beta(y_{t-1} - y_{t-2}) + \varrho\left(\frac{I_t - I_{t-5}}{I_{t-5}}\right) + \varepsilon_t + \vartheta\varepsilon_{t-12}$
Activity univariate ARIMA augmented by Energy extended with…	
BCI	$y_t - y_{t-1} = \gamma(X_t - X_{t-1}) + \delta(X_{t-1} - X_{t-2}) + \kappa(I_t - I_{t-4}) + \varepsilon_t$
EPI	$y_t - y_{t-1} = \gamma(X_t - X_{t-1}) + \delta(X_{t-1} - X_{t-2}) + \varsigma(I_t - I_{t-5}) + \varepsilon_t$
Shock : $\varepsilon_t = \rho\varepsilon_{t-1} + \upsilon_t - \alpha\upsilon_{t-12}$	

Source: Author's elaboration.

In the next section we report the results of predictive exercises both in-sample and out-of-sample.

VI. Forecasting Exercises

1. In-sample results

Table 4 shows some statistics from regressions between activity, a constant, the BCI or EPI contemporaneous and 12 of his lags. In essence, this regression corresponds to the estimated ADL models aforementioned. We report the F-statistic of joint significance of the regression, the p-value of this test and the coefficient of determination of the regression (R^2).

Table 4. Predictive capacity of BCI and EPI for Activity. ADL Model

Variable	F-statistic	P-value	R^2
BCI	7.73	0.000	0.68
EPI	1.39	0.176	0.34
Energy	7.55	0.000	0.47

Source: Author's elaboration.

In the absence of other explanatory variables, the BCI have capacity to predict activity. On the other hand, the EPI doesn’t show capacity to predict activity, we can see this due to F-statistic of joint significance can’t reject the null hypothesis of absence of predictability. Attracts attention the coefficient of determination of BCI which is close to 70%.

Table 5 report the result of the Granger causality analysis to test predictive capacity of BCI and EPI on activity, we find out that BCI and Energy have in-sample predictive capacity, we can see this due to F-statistic of joint significance reject the null hypothesis of absence of predictability, this does not occur with EPI. Also BCI deliver the lowest prediction root mean square error (RMSE). These results are important and complementary to those of the table 4, because now assessed the additional predictive ability of the lags of BCI, EPI and Energy above that which provides the lags of the activity. As well, we see that predictive gains of introducing the BCI are better than the EPI and Energy, allowing a decrease of RMSE between 30%, 13% and 21%, respectively.

Table 5. Granger Causality analysis: predictive capacity of BCI and EPI on annual variation of activity.

Variable	Test F	Degrees of Freedom	P-value	Chi-squared Test	Degrees of Freedom	P-value	RMSE	Quotient RMSE	AIC
BCI	3.40	(12, 97)	0.000	40.73	12	0.000	0.80	0.70	Down
EPI	1.22	(12, 97)	0.280	14.64	12	0.261	0.99	0.87	UP
Energy	2.02	(12,97)	0.029	24.33	12	0.018	0.91	0.79	Down

Source: Author's elaboration.

Table 6 is very similar to table 5, but instead of relying on a VAR specification, is based on the SARIMA specification of annual variation of activity reported in table 2. We estimate three SARIMA specifications augmented by contemporaneous BCI or EPI, then contemporaneous BCI or EPI with 6 of their lags, and finally contemporaneous BCI or EPI with 12 of their lags. Table 6 shows the results of t and F test on related terms of BCI or EPI. With this exercise we look to prove whether business confidence or consumer confidence indicators have predictive information additional to that provides a good SARIMA specification for dependent variable.

Table 6. Granger causality in-sample: predictability of BCI or EPI on annual variation of activity SARIMA specifications

Variable	Test t (1 parameter)	P-value	Test F (7 parameters)	P-value	Test F (13 parameters)	P-value
BCI	-2.44	0.016	5.70	0.000	3.00	0.001
EPI	1.63	0.105	1.66	0.124	2.00	0.027
Energy	6.20	0.000	5.04	0.000	4.67	0.000

Source: Author's elaboration.

Table 6 shows better result of predictability for BCI and Energy than EPI. For one parameter both indicators, BCI and EPI, do not show statistical significance, Energy does. One possible interpretation of this phenomenon is related to the weak predictive contribution of BCI or EPI level in relation to good predictive capability that delivers a good SARIMA specification, i.e. its lags and the evolution of these. Finally, for BCI, EPI and Energy as lags are added predictability increases, but for EPI is statistical significance when we add 12 lags, instead of BCI and Energy which is statistical significance at 6 lags.

2. Out-of-sample results

Predictability analyses reported in table 4 to 6 are often criticized on the basis of strong overfitting of the data. Overall, the in-sample predictability analyses tend to find “false positives” quite often. That is, the in-sample analysis tends to find more predictability than actually there are (see Stock and Watson, 2003). For this reason, that literature is recommended to complete the studies "in sample" with so-called "out of sample". The idea of these last exercises is to estimate predictive models in a portion of the sample, generate forecasts, and then evaluate them using the rest of the sample that has not been considered in the estimation process. Doing recursively these exercises we can construct out-of-sample forecast that, in principle, are not contaminated (or to a lesser extent) by overfitting problems.

Following the recursive methodology described in Section IV, our first out-of-sample analysis is analogous to that shown in table 6, but now using the Clark and West (2007) test describe in the same Section. This is simply a test of out-of-sample statistical significance for different augmented models with BCI or EPI in a context of nested models, because of that we have to do the exercises for the 3 benchmark[10].

Even though we have DMW test, the most important results are those that come from the CW test, if we reject the null hypothesis we know it will be a linear combination (a weighted average) of these two forecast that will generate the lowest MSPE (Clark and West, 2007; Pincheira, 2012).

The table 7 shows the results of CW test for BCI and EPI at different horizons in the context of the first benchmark, the naïve. We can see that BCI is statistically significant at 1% for horizons 2 (nowcast) to 9, and at 5% for horizons 1 (backcast) and 12. For EPI is not statistically significant at any level or horizon. This means that under a naïve model the

[10] The results for different augmented models by BCI or EPI are available upon request. These results are for recursive window, to the results with rolling window see appendix.

addition of BCI into the model is statistically significant at different levels and for different horizons, adding BCI improves forecasts, and we cannot say the same with EPI.

Table 7. Granger Causality out-of-sample: Clark and West (2007) test first benchmark " naïve"

forecast horizons, in months	BCI		EPI	
	CW_t	P-value CW*	CW_t	P-value CW*
1	1.072	0.059	-1.859	0.929
2	2.157	0.010	-1.430	0.805
3	2.346	0.009	-1.100	0.698
6	2.460	0.006	0.145	0.348
9	2.493	0.012	0.088	0.410
12	1.849	0.045	-0.268	0.482
18	1.400	0.112	-0.917	0.667
24	1.367	0.120	-1.196	0.746

Source: Author's elaboration.
Note: recursive window
*P-value obtained from Monte Carlo simulations

Table 8 shows the results of CW test for BCI and EPI at different horizons in the context of the second benchmark, the univariate ARIMA. For estimation problems aforementioned, we not use a SARIMA specification strictly speaking, but incorporate the lag number twelve of the shock, i.e., we do not incorporate the multiplicative effect.

Table 8. Granger Causality out-of-sample: Clark and West (2007) test univariate ARIMA benchmark

forecast horizons, in months	BCI		EPI	
	CW_t	P-value CW*	CW_t	P-value CW*
1	1.825	0.007	-0.772	0.589
2	1.894	0.012	-0.368	0.429
3	2.017	0.018	-0.164	0.352
6	1.837	0.029	0.512	0.191
9	1.871	0.025	1.608	0.050
12	2.095	0.024	1.764	0.045
18	-0.541	0.537	0.410	0.339
24	-1.524	0.816	-1.955	0.908

Source: Author's elaboration.
Note: Recursive window
*P-value obtained from Monte Carlo simulations

We can see that BCI is statistically significant at 1% for horizons 1 and 2 and at 5% for horizons 3 to 12. EPI is statistically significant at 5% for the horizon 9 and 12, only. Is strange to think that if at horizons closest the forecasts are bad, then when the horizon increases the forecasts improve, this is why these results should be taken with caution.

Table 9 shows results for the CW test for BCI and EPI at different horizons in the context of the third benchmark, energy augmented. Again we do not use a SARIMA specification strictly speaking. By expanding the univariate model with energy data and use it as a benchmark, the statistical significance of EPI is found for horizons 9 and 12 only (medium term), while for BCI is significant at 1% for the backcast, at 5% for the nowcast and the first forecast (three months ahead) and at 10% for horizons 6 to 12. Again we find that BCI improves the backcast, nowcast and some forecast of activity under a more complex and complete model.

Table 9. Granger Causality out-of-sample: Clark and West (2007) test energy augmented benchmark

forecast horizons, in months	BCI		EPI	
	CW_t	P-value CW*	CW_t	P-value CW*
1	2.229	0.004	-0.044	0.334
2	2.296	0.013	-0.302	0.440
3	2.016	0.027	-0.464	0.511
6	1.628	0.059	0.146	0.360
9	1.640	0.068	1.326	0.102
12	1.683	0.069	1.661	0.074
18	-0.869	0.645	0.019	0.458
24	-1.561	0.852	-1.875	0.918

Source: Author's elaboration.
Note: Recursive window
*P-value obtained from Monte Carlo simulations

From these three tables we can see that as the benchmark is improved, the results for BCI are maintained, for EPI we find predictive ability in out-of-sample exercises in the medium term, but again we take the results with caution. Remember that if we can reject the null hypothesis of CW test, we know that is going to exist a linear combination of the two forecasts that generate the lowest MSPE. For this reason, we place further attention to this test.

Tables 15-18 (in appendix) show simulated critical values for different significance levels and horizons, it is noteworthy that the simulated values generally are smaller than normal critical values, proposed by Clark and West (2007) for this test. Especially for the first forecast horizon, where direct and iterative method are the same, this can be justified by the size of the sample we are using, because the normal critical values are asymptotic i.e. when the sample is large. To corroborate this, we increase forecast window (and estimation window also), doing that the simulated critical values approach to approximately normal critical values[11].

As robustness check, we estimate the CW statistic for different intervals of time, to check whether the results are driven by the Great Recession or not. Given the results of above, we test the robustness of BCI in the model augmented by Energy and the first three forecasts (backcast, nowcast and first forecast), we see CW statistic bigger during 2013-2014 and lower in post-crisis period. Even in the period of recession in Chile, we observe a CW statistic bigger than 1, which is close to a significance level of 10% (according to the Monte Carlo simulations)[12]. So we conclude that CW results for BCI are not driven by the Great Recession.

Beyond the results presented in the above tables, it is valid to ask for predictive gains that could be obtained by incorporating the BCI and EPI variables in benchmark specifications for generate forecasts one step ahead and at multiple horizons. The tables below (table 10 to 12) show the results for the three benchmarks aforementioned, they are result of a predictive out-of-sample evaluation comparing nested models. We make this comparison for different horizons and use both measure of predictive accuracy, the root mean square prediction error (RMSPE) and the mean absolute prediction error (MAPE). Also we show the P-value of DMW test for both measures[13].

The table 10 shows the result of out-of-sample predictive accuracy using the naïve benchmark, we can see that BCI has RMSPE ratio and MAPE ratio lower than one, that suggest that the augmented model is best than the benchmark, for horizons 2 to 24. But the

[11] These results are available upon request.
[12] See appendix for further explanation.
[13] The results for different augmented models by BCI or EPI are available upon request. These results are for recursive window, to the results with rolling window see appendix.

difference is statistically significant at 10% only for horizons 2 to 12 for RMSPE and horizons 6 to 12 for MAPE. For EPI the results are not as good as expected, the MAPE ratio is lower than one for horizons 6 to 18, but the difference is no statistically significant. As we mentioned, the first forecast correspond to a “backcast” and the second forecast, and most important, is the “nowcast”. So we can say that BCI deliver a best nowcast that the naïve benchmark.

Table 10: Out-of-sample predictive accuracy using a naïve benchmark

forecast horizons, in months	RMSPE Benchmark only	MAPE Benchmark only	RMSPE Benchmark with …	MAPE Benchmark with …	RMSPE Ratio	MAPE Ratio	P-value DMW MSPE	P-value DMW MAPE
				BCI				
1	1.936	1.365	1.928	1.368	0.995	1.002	0.398	0.539
2	2.187	1.683	2.076	1.631	0.949	0.969	0.067	0.187
3	2.556	1.938	2.333	1.852	0.913	0.955	0.036	0.232
6	3.532	2.584	3.224	2.367	0.913	0.916	0.028	0.029
9	4.374	3.478	4.001	3.220	0.915	0.926	0.020	0.046
12	5.317	4.097	4.937	3.742	0.929	0.913	0.069	0.026
18	5.667	4.222	5.274	3.955	0.931	0.937	0.147	0.130
24	5.886	4.504	5.532	4.162	0.940	0.924	0.165	0.114
				EPI				
1	1.936	1.365	1.952	1.385	1.008	1.014	0.980	0.985
2	2.187	1.683	2.228	1.722	1.019	1.023	0.942	0.961
3	2.556	1.938	2.603	1.978	1.018	1.021	0.898	0.933
6	3.532	2.584	3.537	2.551	1.001	0.987	0.536	0.216
9	4.374	3.478	4.383	3.444	1.002	0.990	0.544	0.292
12	5.317	4.097	5.365	4.014	1.009	0.980	0.674	0.157
18	5.667	4.222	5.873	4.192	1.036	0.993	0.866	0.405
24	5.886	4.504	6.263	4.578	1.064	1.017	0.918	0.659

Source: Author's elaboration.

Note: Recursive window

The table 11 shows the result of out-of-sample predictive accuracy using the univariate ARIMA benchmark, we can see that BCI has RMSPE ratio lower than one for horizons 1 to 6, and MAPE ratio lower than one for horizons 2 to 4, that suggest that the augmented model is best than the benchmark, for horizons 1 to 6 and 1 to 9, but the differences are not statistically significance. For EPI the results do not show improvement, the RMSPE and MAPE ratios are lower than one for horizons 9 and 12. The differences are statistically significance for MSPE at 10% only for horizon 12. Even though EPI shows a statistically significant difference we take this with caution, while BCI does not present statistically significant differences at any level.

Table 11: Out-of-sample predictive accuracy using univariate ARIMA benchmark

forecast horizons, in months	RMSPE Benchmark only	MAPE Benchmark only	RMSPE Benchmark with …	MAPE Benchmark with …	RMSPE Ratio	MAPE Ratio	P-value DMW MSPE	P-value DMW MAPE
				BCI				
1	1.399	1.032	1.395	1.077	0.998	1.043	0.480	0.801
2	1.650	1.242	1.598	1.236	0.968	0.995	0.348	0.472
3	1.932	1.509	1.822	1.413	0.943	0.936	0.262	0.245
6	2.651	2.074	2.553	2.127	0.963	1.025	0.371	0.585
9	3.263	2.439	3.446	2.776	1.056	1.138	0.640	0.803
12	3.615	2.659	4.363	3.384	1.207	1.273	0.797	0.878
18	2.883	2.305	5.342	3.829	1.853	1.662	0.935	0.983
24	2.643	2.091	5.980	4.229	2.263	2.022	0.968	0.995
				EPI				
1	1.399	1.032	1.447	1.099	1.035	1.065	0.956	0.985
2	1.650	1.242	1.728	1.321	1.047	1.063	0.906	0.955
3	1.932	1.509	2.038	1.560	1.055	1.034	0.861	0.754
6	2.651	2.074	2.699	2.088	1.018	1.007	0.626	0.541
9	3.263	2.439	3.044	2.384	0.933	0.978	0.131	0.359
12	3.615	2.659	3.263	2.531	0.902	0.952	0.078	0.257
18	2.883	2.305	2.987	2.401	1.036	1.042	0.666	0.707
24	2.643	2.091	3.246	2.593	1.228	1.240	0.996	0.994

Source: Author's elaboration.

Note: Recursive window

The table 12 shows the result of out-of-sample predictive accuracy using energy augmented benchmark. In presence of a more accurate model the results for BCI are less promising; we see that the RMSPE and MAPE ratio are lower than for horizons 2 to 9 and 3 to 6, respectively, but the differences are not statistically significance. The result for EPI do not change significant, the RMSPE and MAPE ratio are lower than one for horizons 9 and 12 and 12, respectively. Again we can see that the difference is statistically significant for horizon 12. We can say that in presence of a good benchmark the predictive accuracy of BCI loses power to enhance forecast of activity.

Table 12: Out-of-sample predictive accuracy using energy augmented benchmark

forecast horizons, in months	RMSPE Benchmark only	MAPE Benchmark only	RMSPE Benchmark with …	MAPE Benchmark with …	RMSPE Ratio	MAPE Ratio	P-value DMW MSPE	P-value DMW MAPE
				BCI				
1	1.043	0.807	1.064	0.865	1.020	1.072	0.670	0.894
2	1.575	1.229	1.542	1.234	0.979	1.004	0.320	0.528
3	1.899	1.514	1.776	1.433	0.935	0.946	0.146	0.224
6	2.636	1.994	2.466	1.944	0.935	0.975	0.162	0.363
9	3.224	2.430	3.161	2.490	0.980	1.024	0.399	0.615
12	3.537	2.655	3.563	2.768	1.008	1.043	0.531	0.651
18	2.855	2.248	3.787	2.685	1.326	1.194	0.912	0.867
24	2.585	2.007	4.014	2.826	1.553	1.409	0.956	0.969
				EPI				
1	1.043	0.807	1.142	0.880	1.095	1.090	0.946	0.932
2	1.575	1.229	1.724	1.355	1.095	1.103	0.965	0.972
3	1.899	1.514	2.106	1.639	1.109	1.083	0.965	0.918
6	2.636	1.994	2.809	2.134	1.066	1.070	0.907	0.846
9	3.224	2.430	3.171	2.484	0.984	1.022	0.345	0.651
12	3.537	2.655	3.238	2.559	0.916	0.964	0.094	0.284
18	2.855	2.248	3.015	2.368	1.056	1.053	0.768	0.759
24	2.585	2.007	3.138	2.454	1.214	1.223	0.991	0.994

Source: Author's elaboration.
Note: Recursive window

From the 3 tables above, we can see that the accuracy of the benchmark improves as autoregressive and moving average components are added, also when we added the energy variable the forecast improve a lot, for example for the “backcast” it goes from a RMSPE of 1.95 to 1.04 percentage points.

As we mentioned before, we use another loss function, the hit rate, the rate of correctly forecasting the direction of change in activity. Table 13 displays the average hit rate across models and horizons within benchmark and augmented models[14]. Percentages in bold indicate superiority of models according to DMW test. The results are mixed but consistent

[14] These results are for recursive window, to the results with rolling window see appendix.

with those obtained in terms of RMSPE and Clark and West test. At short horizons, BCI augmented models outperform benchmark models in forecasting the direction of change. At longer horizons, we find that EPI augmented models are better off than benchmark models, but the difference is not statistical significance. In general terms, BCI and EPI augmented models outperform naïve benchmark in most of the horizons (6 and 7 of 8 horizons, respectively).

Table 13. Average hit rate across models and horizons

Variable	Model	h=1	h=2	h=3	h=6	h=9	h=12	h=18	h=24
BCI	naïve benchmark	51%	49%	43%	43%	29%	26%	15%	18%
	Augmented	51%	57%	**68%**	**63%**	**55%**	**53%**	**49%**	**46%**
	ARIMA benchmark	65%	69%	72%	58%	75%	73%	**70%**	**85%**
	Augmented	70%	64%	61%	65%	67%	56%	52%	44%
	Energy benchmark	77%	69%	72%	61%	76%	66%	75%	84%
	Augmented	79%	64%	57%	71%	79%	67%	75%	72%
EPI	naïve benchmark	51%	49%	43%	43%	29%	26%	15%	18%
	Augmented	44%	46%	**67%**	**56%**	**46%**	38%	**36%**	21%
	ARIMA benchmark	65%	69%	72%	58%	75%	73%	70%	85%
	Augmented	65%	70%	61%	57%	72%	77%	75%	72%
	Energy benchmark	77%	69%	72%	61%	76%	66%	75%	84%
	Augmented	76%	64%	63%	61%	74%	68%	75%	75%

Source: Author's elaboration

Note: Recursive window of estimation

The results of forecasting exercises show mixed results. In detail, from CW test we can conclude that EPI is not statistically significant for the backcast and nowcast in the 3 benchmarks for nearby horizons, we see some predictive ability for medium term forecast, 9 and 12 months ahead. In the other hand, BCI is statistically significant for the backcast, nowcast and some forecast in the 3 benchmarks, so we can say that is going to exist a linear combination between the model with energy and the model with energy and BCI that gives the lowest MSPE, independent that DMW test do not find that the model with energy and BCI is better than the model with energy only. The results from the hit rate exercises are consistent with those found with the DMW test. So, we conclude that the business confidence indicator can be use as a leading indicator of the Chilean activity.

VII. Conclusion

In this article we evaluate, in several dimensions, the ability of two monthly sentiment indicators to forecast activity in Chile. We place special attention to the role that these sentiment indicators may be playing to improve the accuracy of activity backcasts and nowcasts. We do in-sample and out-of-sample exercises to test whether BCI or EPI have predictive ability. We place attention in out-of-sample over in-sample exercises because is well known that the latter suffer from overfitting.

In the out-of-sample analysis, we find that the addition of BCI in the naïve benchmark is statistically significant, according to the CW test. In this context, the DMW result shows that BCI improve the predictive accuracy of the naïve model, for the EPI we do not find predictability gains. For the second benchmark, the addition of BCI is statistically significant in short and medium term forecast and the addition of EPI is statistically significant only in medium term forecast (nine and twelve months ahead), according to the CW test. Using the second benchmark, the DMW results do not show predictability gains for the two sentiment indicators.

For the third benchmark, DMW test say that the model augmented by energy and BCI is not better, in predictive accuracy, that model augmented by energy only. But the CW results shows that BCI is statistically significant in the out-of-sample exercises, so is going to exist a linear combination between the two nested models that deliver a lowest MSPE, we cannot say the same for EPI in short term forecast, only in medium term forecast (nine and twelve months ahead). The robustness check shows that for BCI the results of CW test are not driven by the Great Recession.

In the second and third benchmark, we find that the EPI is statistically significant in medium term forecast, but not in short term forecast, is rare to find predictability in the medium term and not in the short term, so future research could investigate the economic explanation behind this results.

From the Monte Carlo simulations, we see that approximately normal critical values do not yield a test with correct size in our framework, so as Clark and West (2007) recommend, we use the Monte Carlo simulations to find critical values that yield a test with correct size given a certain significance level, i.e. the p-value of our CW statistic. Even though iterative and direct methods are different, the first forecast, one month ahead, are equal under both methods, our simulations for that horizon show critical values lower than the normal values proposed by Clark and West (2007), we suspect that this happen because our sample is not big enough. In our simulations, we expand the numbers of observations and we obtain approximately normal values, proving our hypothesis.

The results of hit rate exercises, the rate of correctly forecasting the direction of change in activity, are consistent with those found with the DMW test, BCI and EPI are more accurate in forecasting the direction of change compared to a naïve benchmark.

Finally, in this paper we prove that the business confidence indicator can be use as a leading indicator of Chilean activity, given that improves backcasts, nowcasts and forecasts of activity. Future researches can study what are the transmission channels and how stable are the results in different economic fluctuations, recessions and booms.

Reference

Box, G. & Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. San Francisco, CA: Holden-Day.

Bravo, F. & Franken, H. (2001). Un indicador Líder del IMACEC. Working paper N°99, Central Bank of Chile.

Calvo, G. & Ricaurte, M. (2012). Indicadores Sintéticos para la Proyección de IMACEC en Chile. Working paper N°656, Central Bank of Chile.

Chauvet, M., & Guo, J. T. (2003). Sunspots, Animal Spirits, and Economic Fluctuations. Macroeconomic Dynamics, 7(1), 140-169.

Chauvet, M., & Potter, S. (2013). Forecasting Output. Handbook of Economic Forecasting, 2(Part A), 141-194.

Christiansen, C., Eriksen, J. & Møller, S. (2013).Forecasting US Recessions: The Role of Sentiment. Research paper, Center for Research in Econometric Analysis of Time Series, Denmark. Available at SSRN: http://ssrn.com/abstract=2217403

Clark, T. & West, K. (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. Journal of Econometrics 138: 291–311.

Diebold, F. & Mariano, R. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics 13(3): 253–63.

Echavarría, G., & González, W. (2011). Un Modelo de Factores Dinámicos de Pequeña Escala para el Imacec. Notas de Investigación Revista Economía Chilena (The Chilean Economy), 14(2), 109-118.

Firinguetti, L. & Rubio, H. (2003). Indicadores Líderes del IMACEC. Working paper N°208, Central Bank of Chile.

Hengelbrock, J., Theissen, E. & Westheide, C. (2011). Market response to investor sentiment. CFS Working Paper, No. 2011/02, http://nbn-resolving.de/urn:nbn:de:hebis:30-91456

Huang, H., Lee, T. H., & Li, C. (2006). Forecasting output growth and inflation: How to use information in the yield curve. mimeo, University of California, Riverside.

Hüfner, F. & Schröder, M. (2002). Forecasting economic activity in Germany: how useful are sentiment indicators?. ZEW Discussion Papers 02-56, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.

Ludvigson, S. C. (2004). Consumer Confidence and Consumer Spending. Journal of Economic Perspectives, 18(2), 29-50.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. Journal of Econometrics, 135(1), 499-526.

Pedersen, M. (2009). Un indicador líder compuesto para la actividad económica en Chile. Revista Monetaria, vol. XXXII, No. 2 april-june, CEMLA.

Pesaran, M. H., & Weale, M. (2006). Survey Expectations. Handbook of Economic Forecasting, 1, 715-776.

Pincheira, P. (2012). Are Forecast Combinations Efficient?. Working Papers Central Bank of Chile 661, Central Bank of Chile.

Pincheira, P. (2014). Predicción del Empleo Sectorial y Total en Base a Indicadores de Confianza Empresarial. Revista Economía Chilena, vol. 17, No. 1, Banco Central de Chile.

Pošta, V. & Pikhart, Z. (2012). The Use of the Sentiment Economic Indicator for GDP Forecasting: Evidence from EU Economies. In Statistika, Economy and Statistics Journal, 2012, vol. 1, pp. 41-55.

Stock, J. H. & Watson, M. W. (2002). Macroeconomic Forecasting Using Diffusion Indexes. Journal of Business & Economic Statistics 20(2): 147–62.

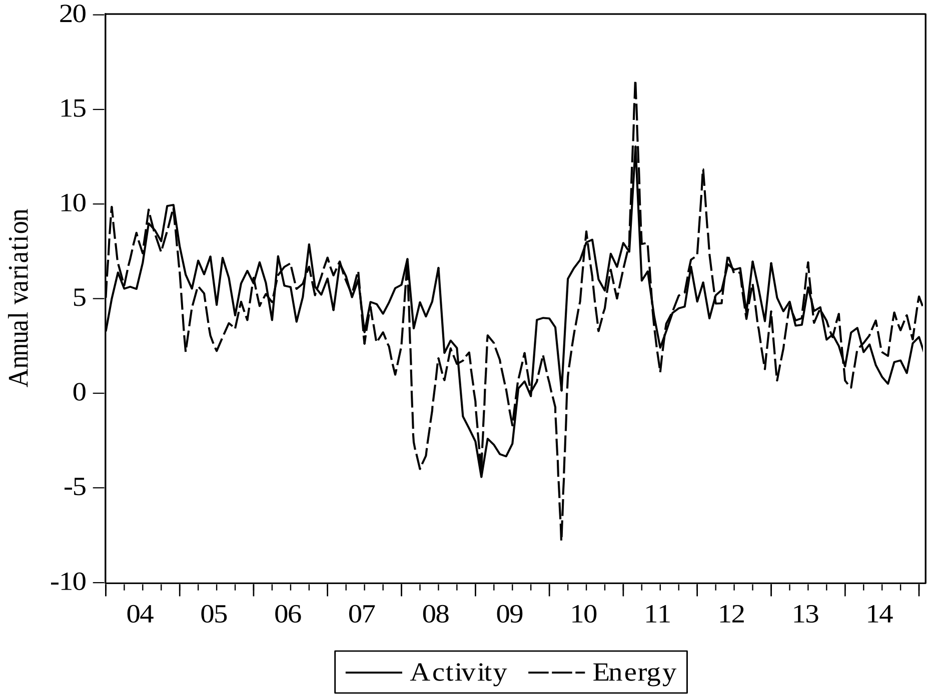
Stock, J. H. & Watson, M. W. (2003). Forecasting Output and Inflation: The Role of Asset Prices. Journal of Economic Literature, 41(3): 788-829.

Urrutia, M. & Sánchez, A. (2008). Generación eléctrica en un modelo para proyectar el IMACEC. Revista Economía Chilena, vol. 11, No. 2, Central Bank of Chile.

West, K. (1996). Asymptotic Inference about Predictive Ability. Econometrica 64(5):1067–84.

Appendix

Figure 1A. Activity and Energy,
January 2004 - February 2015



Source: Author's elaboration, Central Bank of Chile, Center for Economic Load Dispatch.

Hit Rate exercises:

The hit rate exercise is to assess whether the model is able to forecast the change of direction of the variable of interest. To do this we proceed as follows: we take the actual data of the month we want to forecasting (y_t) then we subtract the last data we know at the time of making the forecast (y_{t-1}), multiply this by the forecast we make (y_t^f) less the latest date (y_{t-1}), if this is greater or equal to zero takes value zero if less than zero takes value one. We do this for all the periods we do forecast and them take an average, the average corresponds to the failure rate; one minus the failure rate will be the hit rate.

$$(y_t - y_{t-1}) * \left(y_t^f - y_{t-1}\right) \; si \begin{cases} \geq 0 \; take \; value \; 0 \\ < 0 \; take \; value \; 1 \end{cases}$$

Monte Carlo simulation:

For the out-of-sample exercises, we use Clark and West (2007) test, this test works well in direct method of forecast, in this paper we use iterative method of forecast, so the normal critical values that the authors propose might not render the correct size. To deal with this problem we resort to Monte Carlo simulations. The following explains the procedure:

We generate the 4 variables uses in this paper (Activity, BIC, EPI and Energy), the data generate process (DGP) shown below[15], we generate 600 observations for each series and take the last 134 observations, we do this because, in general, last generated observations tend to be more stationary than the first. Then with the final observations we recreate the out-of-sample exercises explain in section IV.

Table 14. Data generating process for Monte Carlo simulations:

Activity	$y_t = 0.6 + 0.85 * y_{t-1} + \varepsilon_t - 0.9 * \varepsilon_{t-12} \qquad \varepsilon_t \sim (0,1.2)$
Activity generated by energy and lags	$y_t = 0.4 + 0.83 * y_{t-1} + 0.45 * x_t - 0.37 * x_{t-1} + \varepsilon_t - 0.91 * \varepsilon_{t-12} \qquad \varepsilon_t \sim (0,0.9)$
BCI	$I_t = 2.6 + 0.95 * I_{t-1} + \epsilon_t + 0.21 * \epsilon_{t-12} \qquad \epsilon_t \sim (0,2.2)$
EPI	$I_t = 3.9 + 0.92 * I_{t-1} + \epsilon_t + 0.09 * \epsilon_{t-12} \qquad \epsilon_t \sim (0,2.4)$
Energy	$x_t = 0.9 + 0.74 * X_{t-1} + u_t - 0.93 * u_{t-12} \qquad u_t \sim (0,1.7)$

Source: Author's elaboration.

Repeat the above 5,000 times, in every repetition we keep the CW t statistic for each horizon. When the routine ends, we sort CW t statistic from low to high and seek in which part of the distributions is located the CW t statistic that comes from the original data. Then the p-value will be the percentage of statistics that remain below the original statistical.

[15] For activity and energy the DGP are the models with lowest Akaike, for BCI and EPI are parsimonious models with good coefficient of determination and statistically significant parameters.

In the DGP, we establish the null hypothesis of not predictability, i.e. BCI and EPI are not part of the generative process of activity. Notice that we have to do this simulation for each benchmark, variable (BCI and EPI) and method of estimation (recursive and rolling window).

Simulated Critical Values:

Table 15: Simulated critical values of Clark and West (2007) test for BCI at different horizons in recursive window of estimation

BCI	Naïve benchmark			ARIMA benchmark			Energy benchmark		
	Significance level (α)			Significance level (α)			Significance level (α)		
H=	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	0.797	1.159	1.915	0.722	1.107	1.908	0.757	1.142	1.874
2	0.910	1.289	2.133	0.822	1.258	2.130	0.903	1.300	2.167
3	1.003	1.434	2.292	0.948	1.393	2.226	1.025	1.459	2.222
6	1.070	1.476	2.240	1.044	1.501	2.289	1.179	1.634	2.393
9	1.276	1.731	2.534	1.139	1.573	2.328	1.236	1.669	2.331
12	1.342	1.796	2.629	1.277	1.726	2.502	1.344	1.765	2.491
18	1.481	1.984	2.806	1.568	1.978	2.716	1.559	1.993	2.778
24	1.515	1.930	2.690	1.601	1.986	2.741	1.588	1.993	2.763
Normal	1.280	1.645	2.330	1.280	1.645	2.330	1.280	1.645	2.330

Source: Author's elaboration

Table 16: Simulated critical values of Clark and West (2007) test for EPI at different horizons in recursive window of estimation

EPI	Naïve benchmark			ARIMA benchmark			Energy benchmark		
	Significance level (α)			Significance level (α)			Significance level (α)		
H=	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	0.874	1.250	1.968	0.776	1.164	1.983	0.795	1.173	1.791
2	1.030	1.432	2.168	0.902	1.318	2.257	0.927	1.356	2.106
3	1.134	1.568	2.322	1.025	1.481	2.349	1.077	1.518	2.270
6	1.223	1.691	2.519	1.111	1.553	2.370	1.241	1.630	2.433
9	1.393	1.800	2.573	1.174	1.592	2.362	1.248	1.625	2.326
12	1.375	1.792	2.610	1.320	1.746	2.514	1.324	1.750	2.457
18	1.515	1.941	2.709	1.582	2.021	2.731	1.561	1.996	2.762
24	1.553	2.003	2.729	1.627	2.010	2.775	1.625	2.027	2.678
Normal	1.280	1.645	2.330	1.280	1.645	2.330	1.280	1.645	2.330

Source: Author's elaboration

Table 17: Simulated critical values of Clark and West (2007) test for BCI at different horizons in rolling window of estimation

BCI	Naïve benchmark			ARIMA benchmark			Energy benchmark		
	Significance level (α)			Significance level (α)			Significance level (α)		
H=	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	0.784	1.144	1.865	0.750	1.084	1.753	0.799	1.119	1.731
2	0.927	1.279	1.984	0.836	1.176	1.957	0.866	1.208	1.994
3	0.990	1.406	2.227	0.936	1.341	2.108	0.968	1.363	2.124
6	1.143	1.549	2.258	1.086	1.514	2.228	1.162	1.549	2.275
9	1.360	1.820	2.588	1.176	1.633	2.337	1.264	1.708	2.477
12	1.409	1.878	2.662	1.187	1.685	2.509	1.305	1.720	2.341
18	1.609	2.041	2.757	1.511	1.927	2.645	1.560	1.923	2.678
24	1.586	2.006	2.741	1.569	1.937	2.635	1.583	1.941	2.599
Normal	1.280	1.645	2.330	1.280	1.645	2.330	1.280	1.645	2.330

Source: Author's elaboration

Table 18: Simulated critical values of Clark and West (2007) test for EPI at different horizons in rolling window of estimation

EPI	Naïve benchmark			ARIMA benchmark			Energy benchmark		
	Significance level (α)			Significance level (α)			Significance level (α)		
H=	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	0.896	1.240	1.861	0.791	1.181	1.881	0.817	1.131	1.843
2	1.062	1.435	2.136	0.947	1.286	2.113	0.920	1.294	1.975
3	1.169	1.582	2.323	1.066	1.439	2.179	1.071	1.454	2.128
6	1.320	1.719	2.498	1.157	1.571	2.360	1.210	1.585	2.264
9	1.447	1.848	2.592	1.245	1.674	2.501	1.299	1.685	2.414
12	1.437	1.877	2.639	1.265	1.733	2.508	1.344	1.700	2.343
18	1.645	2.081	2.778	1.504	1.921	2.664	1.580	1.927	2.549
24	1.643	2.087	2.811	1.560	1.914	2.634	1.580	1.935	2.525
Normal	1.280	1.645	2.330	1.280	1.645	2.330	1.280	1.645	2.330

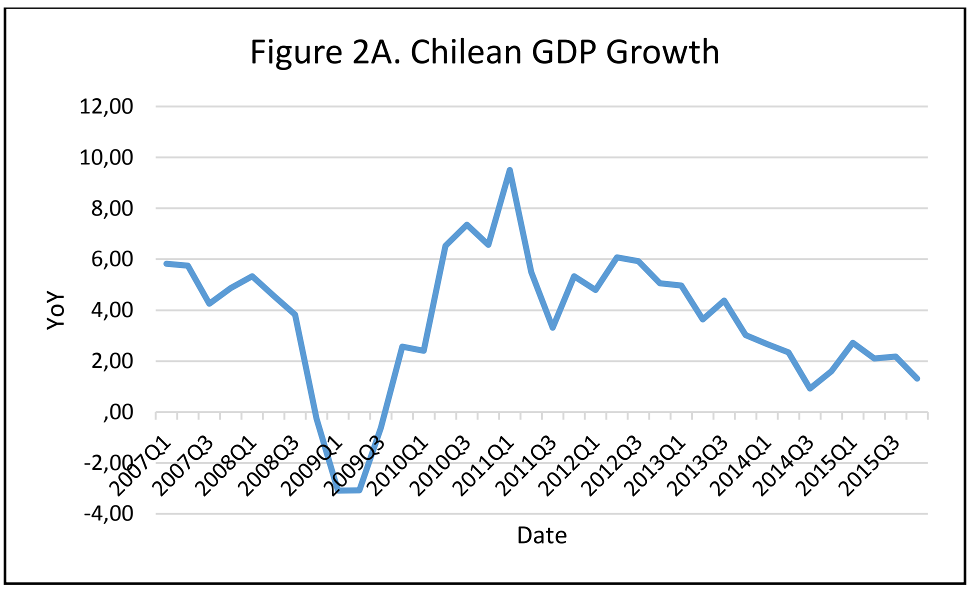
Source: Author's elaboration

Robustness check

In this section we check the robustness of results of the section VI. If we look up the figure 1 again, one can think that the strong relationship between BCI and activity occur mainly during the Great Recession, and like crisis are atypical events the relationship between BCI and activity during normal times could be less strong. To prove that the latter is not the case for our results, we test the Clark and West test for different intervals of time.

The results we show in the section VI for the CW test are for the whole sample, this is, we compute the CW statistic from all errors of forecast, so errors of forecast in the great recession are included in the CW results. To avoid this, we do the following exercise, we compute the CW test using a rolling window of size R=25, for example, we have 84 error of forecast for the first horizon (the backcast), we compute the CW statistic using the first 25 error of forecast, then we move the rolling window to include the following 25 observations, so we eliminate the first and include the 26th error of forecast, for this second rolling window we compute the CW statistic, we do this up to the last error of forecast. Finally, we have 59 CW statistic, which cover different intervals of time.

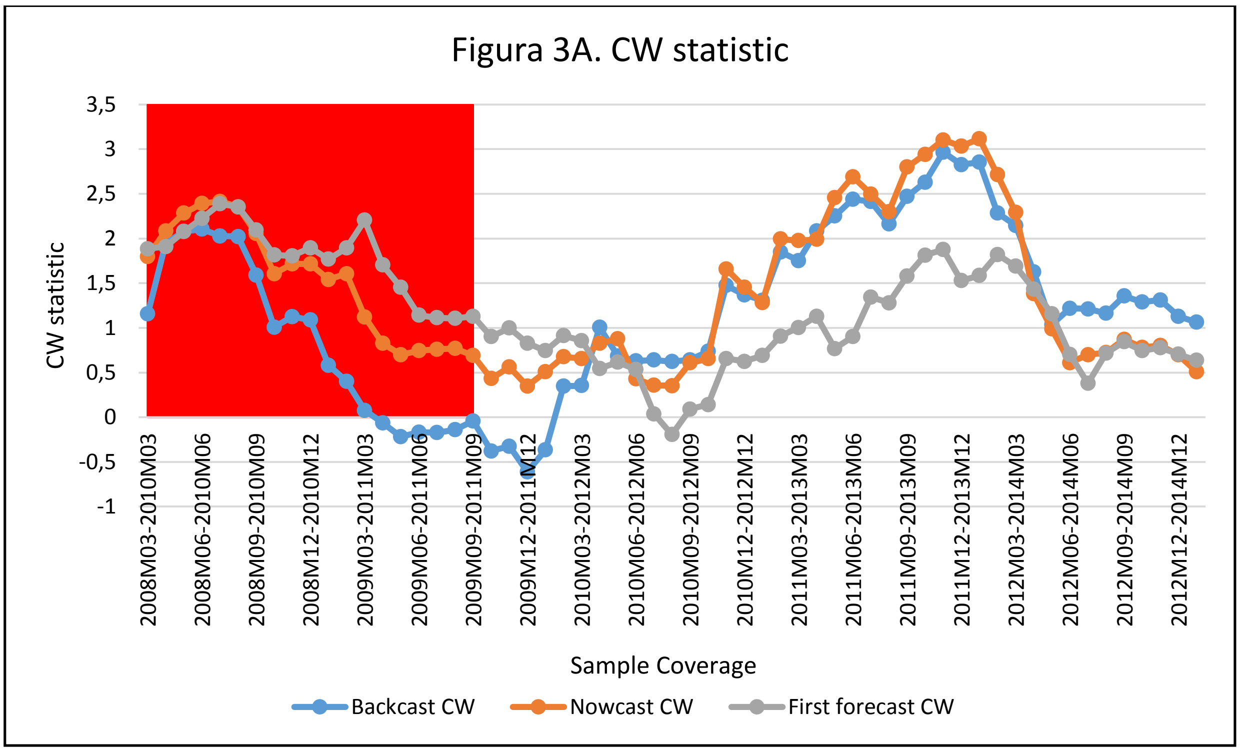
To define the period of recession we use the practical definition[16], two consecutive quarters of negative growth. So using Chile GDP growth data, we define that the Great Recession lasted from 2008M10 to 2009M09, like it shows in the figure 2A.



Source: Author’s elaboration, Central Bank of Chile.

[16] See for example http://news.bbc.co.uk/2/hi/business/7495340.stm or http://www.imf.org/external/pubs/ft/fandd/2009/03/basics.htm

Given the results of section VI, we focus in the three first forecast (backcast, nowcast and first forecast) and the BCI variable, using the model augmented by Energy[17]. If the results of section VI are strong only given the crisis we should see a CW statistic bigger during the crisis and smaller in normal times. The figure 3A shows the CW statistic for the 3 horizons and the model aforementioned, the area marked in red corresponds to windows where there are still observations that belong to the crisis in Chile. The figure shows that the results of section VI are not driven by the crisis, in fact the statistical significance of BCI is less clear during the post-crisis and strong in normal times, remember that a CW bigger of 1 is close to a significance level of 10%, using the Monte Carlo simulations.



Source: Author's elaboration
Note: The red area corresponds to windows where there are still observations that belong to the crisis in Chile. The dots of lines correspond to the CW statistic for every rolling window estimation.

[17] The results for EPI and the others models are available upon request.

The results show here demonstrate that BCI is statistical significance during normal times and the relationship is less clear during post-turmoil time. Is noteworthy, that the backcast had a poorly performing during the recession, and his performance has improved in the last years (2013-2014). Finally, we can conclude that our results are not only explain by the Great Recession.

Out-of-sample results using rolling window:

Table 7a. Granger Causality out-of-sample: Clark and West (2007) test first benchmark "naive"

forecast horizons, in months	BCI		EPI	
	CW_t	P-value CW*	CW_t	P-value CW*
1	0.919	0.079	-0.874	0.662
2	2.127	0.009	-0.512	0.526
3	2.298	0.008	0.154	0.320
6	2.307	0.009	2.025	0.028
9	2.333	0.018	1.902	0.044
12	2.111	0.032	2.052	0.035
18	2.032	0.051	1.687	0.095
24	2.067	0.044	2.217	0.039

Source: Author's elaboration.
Note: Rolling window

*P-value obtained from monte carlo simulations

Table 8a. Granger Causality out-of-sample: Clark and West (2007) test univariate ARIMA benchmark

forecast horizons, in months	BCI		EPI	
	CW_t	P-value CW*	CW_t	P-value CW*
1	1.884	0.008	-0.263	0.423
2	1.817	0.014	-0.839	0.622
3	2.060	0.012	-0.045	0.353
6	1.987	0.018	0.847	0.158
9	2.048	0.021	1.764	0.042
12	2.427	0.012	1.813	0.042
18	-0.174	0.434	0.302	0.348
24	-1.468	0.798	-2.135	0.946

Source: Author's elaboration.
Note: Rolling window

*P-value obtained from Monte Carlo simulations

Table 9a. Granger Causality out-of-sample: Clark and West (2007) test energy augmented benchmark

forecast horizons, in months	BCI		EPI	
	CW_t	P-value CW*	CW_t	P-value CW*
1	0.879	0.089	-0.506	0.523
2	1.402	0.034	-0.182	0.393
3	1.525	0.037	-0.250	0.435
6	1.083	0.112	0.801	0.179
9	1.110	0.121	2.210	0.015
12	1.535	0.069	2.120	0.021
18	-0.750	0.621	0.279	0.377
24	-1.628	0.856	-2.102	0.949

Source: Author's elaboration.
Note: Rolling window

*P-value obtained from Monte Carlo simulations

Table 10a: Out-of-sample predictive accuracy using a naive benchmark

forecast horizons, in months	RMSPE Benchmark only	MAPE Benchmark only	RMSPE Benchmark with …	MAPE Benchmark with …	RMSPE Ratio	MAPE Ratio	P-value DMW MSPE	P-value DMW MAPE
				BCI				
1	1.948	1.374	1.946	1.409	0.999	1.025	0.477	0.891
2	2.221	1.718	2.095	1.674	0.943	0.974	0.061	0.235
3	2.618	2.002	2.354	1.876	0.899	0.937	0.032	0.175
6	3.691	2.774	3.295	2.496	0.893	0.900	0.033	0.037
9	4.672	3.794	4.149	3.385	0.888	0.892	0.022	0.027
12	5.815	4.539	5.145	3.941	0.885	0.868	0.031	0.008
18	6.605	4.883	5.663	4.262	0.857	0.873	0.038	0.036
24	7.048	5.273	5.987	4.402	0.849	0.835	0.036	0.022
				EPI				
1	1.948	1.374	1.968	1.405	1.010	1.022	0.922	0.970
2	2.221	1.718	2.266	1.775	1.020	1.033	0.832	0.943
3	2.618	2.002	2.641	1.997	1.009	0.998	0.630	0.466
6	3.691	2.774	3.530	2.660	0.956	0.959	0.038	0.096
9	4.672	3.794	4.428	3.554	0.948	0.937	0.040	0.034
12	5.815	4.539	5.494	4.218	0.945	0.929	0.023	0.026
18	6.605	4.883	6.261	4.667	0.948	0.956	0.074	0.115

24	7.048	5.273	6.760	4.864	0.959	0.922	0.025	0.009

Source: Author's elaboration.

Note: Rolling window

Table 11a: Out-of-sample predictive accuracy using univariate ARIMA benchmark

forecast horizons, in months	RMSPE Benchmark only	MAPE Benchmark only	RMSPE Benchmark with …	MAPE Benchmark with …	RMSPE Ratio	MAPE Ratio	P-value DMW MSPE	P-value DMW MAPE
				BCI				
1	1.425	1.062	1.405	1.082	0.986	1.019	0.374	0.662
2	1.623	1.254	1.588	1.233	0.978	0.983	0.393	0.400
3	1.938	1.474	1.816	1.433	0.937	0.972	0.251	0.387
6	2.712	2.112	2.563	2.175	0.945	1.030	0.306	0.606
9	3.286	2.458	3.308	2.764	1.006	1.124	0.519	0.792
12	3.645	2.667	4.002	3.148	1.098	1.180	0.692	0.808
18	2.908	2.312	4.751	3.532	1.634	1.528	0.934	0.981
24	2.632	2.090	5.456	4.011	2.073	1.919	0.983	0.998
				EPI				
1	1.425	1.062	1.472	1.123	1.033	1.057	0.952	0.985
2	1.623	1.254	1.745	1.387	1.075	1.106	0.985	0.999
3	1.938	1.474	2.053	1.592	1.060	1.080	0.929	0.953
6	2.712	2.112	2.742	2.162	1.011	1.023	0.584	0.652
9	3.286	2.458	3.093	2.449	0.941	0.996	0.152	0.478
12	3.645	2.667	3.345	2.631	0.918	0.986	0.131	0.440
18	2.908	2.312	3.108	2.540	1.069	1.098	0.743	0.834
24	2.632	2.090	3.441	2.775	1.307	1.328	0.998	0.998

Source: Author's elaboration.

Note: Rolling window

Table 12a: Out-of-sample predictive accuracy using energy augmented benchmark

forecast horizons, in months	RMSPE Benchmark only	MAPE Benchmark only	RMSPE Benchmark with …	MAPE Benchmark with …	RMSPE Ratio	MAPE Ratio	P-value DMW MSPE	P-value DMW MAPE
				BCI				
1	1.047	0.800	1.175	0.946	1.122	1.183	0.993	0.999
2	1.566	1.218	1.631	1.313	1.041	1.078	0.768	0.909
3	1.892	1.505	1.844	1.474	0.975	0.979	0.369	0.393
6	2.639	2.002	2.618	2.107	0.992	1.053	0.462	0.731
9	3.223	2.429	3.379	2.750	1.048	1.132	0.708	0.906
12	3.550	2.643	3.716	2.915	1.047	1.103	0.665	0.798

18	2.870	2.239	3.937	2.754	1.372	1.230	0.919	0.891
24	2.554	1.991	4.260	3.002	1.668	1.508	0.968	0.983
				EPI				
1	1.047	0.800	1.171	0.920	1.119	1.150	0.983	0.997
2	1.566	1.218	1.713	1.358	1.093	1.115	0.983	0.988
3	1.892	1.505	2.085	1.646	1.102	1.094	0.965	0.936
6	2.639	2.002	2.745	2.081	1.040	1.039	0.800	0.718
9	3.223	2.429	3.093	2.392	0.960	0.985	0.142	0.403
12	3.550	2.643	3.141	2.427	0.885	0.918	0.049	0.142
18	2.870	2.239	3.041	2.465	1.059	1.101	0.743	0.871
24	2.554	1.991	3.268	2.613	1.279	1.312	1.000	1.000

Source: Author's elaboration.

Note: Rolling window

Table 13a. Average hit rate across models and horizons

Variable	Model	h=1	h=2	h=3	h=6	h=9	h=12	h=18	h=24
BCI	naïve benchmark	44%	40%	38%	33%	24%	18%	22%	20%
	Augmented	43%	45%	**70%**	**56%**	**39%**	**40%**	**45%**	**48%**
	ARIMA benchmark	64%	67%	73%	59%	76%	73%	75%	**87%**
	Augmented	**71%**	69%	63%	65%	68%	66%	61%	46%
	Energy benchmark	79%	69%	68%	63%	75%	66%	73%	84%
	Augmented	75%	64%	59%	68%	72%	71%	76%	69%
EPI	naïve benchmark	44%	40%	38%	33%	24%	18%	22%	20%
	Augmented	44%	39%	**68%**	41%	**37%**	**30%**	**36%**	**34%**
	ARIMA benchmark	64%	67%	73%	59%	76%	73%	75%	87%
	Augmented	63%	64%	63%	59%	68%	68%	70%	72%
	Energy benchmark	**79%**	69%	68%	63%	75%	66%	73%	84%
	Augmented	70%	67%	65%	65%	79%	**75%**	78%	79%

Source: Author's elaboration

Note: Rolling window of estimation