Can we beat the Random Walk? The case of survey-based exchange rate forecasts in Chile

Pincheira, Pablo and Neumann, Federico

Universidad Adolfo Ibáñez, Universidad Adolfo Ibáñez

8 December 2018

Can we beat the Random Walk?
The case of survey-based exchange rate forecasts in Chile

Pablo Pincheira Brown[♣]

School of Business, Universidad Adolfo Ibáñez

&

Federico Neumann

School of Business, Universidad Adolfo Ibáñez

December 2018

Abstract

We examine the accuracy of survey-based expectations of the Chilean exchange rate relative to the US dollar. Our out-of-sample analysis reveals that survey-based forecasts outperform the Driftless Random Walk (DRW) in terms of Mean Squared Prediction Error at several forecasting horizons. This result holds true even when comparing the survey to a more competitive benchmark based on a refined information set. A similar result is found when precision is measured in terms of Directional Accuracy: survey-based forecasts outperform a "pure luck" benchmark at several forecasting horizons. Differing from the traditional "no predictability" result reported in the literature for many exchange rates, our findings suggest that the Chilean peso is indeed predictable.

Keywords: Survey expectations, Exchange Rates, Forecasting, Random Walk, Directional Accuracy, Mean Squared Prediction Error.

JEL Classification: C01, C32, C52, C53, C58, G17, G11, E270, E370, F370, L740, O180, R310.

[♣] We thank Andrea Bentancor and Nicolás Hardy for wonderful comments.

Pablo Pincheira Brown (corresponding author): Diagonal Las Torres 2640, Peñalolén, Santiago, Chile. Email: pablo.pincheira@uai.cl.

1. Introduction

Our main goal in this paper is to analyze the performance of Chilean exchange rate forecasts coming from the Survey of Professional Forecasters (SPF) carried out by the Central Bank of Chile. As it is widely known, when it comes to exchange rates, the simple Driftless Random Walk (DRW) is a very difficult benchmark to beat in out-of-sample comparisons as shown initially by Meese and Rogoff (1983a, 1983b). Since then, exchange rate predictability has become an obsession in the literature with a number of articles trying to overturn the seminal results of Meese and Rogoff (1983a, 1983b) or simply trying to address the problem from another perspective, using a new dataset, theory or econometric technique. See for instance Cheung, Chinn, and Pascual (2005), Clark and West (2006), Engel and West (2005), Engel, Mark and West (2015), Molodtsova and Papell (2009) and Ince and Molodtsova (2016) just to mention a few. While in the last years, some papers have shown to outperform the DRW, according to the review in Rossi (2013a), "…*Meese and Rogoff's (1983a, 1983b) finding does not seem to be entirely and convincingly overturned*." Rossi (2013a), pages 1113-1114. Given that the DRW is reported in the same review of Rossi (2013a) as the toughest model to beat, we use it as our main benchmark in this paper.

In principle, our focus on the Chilean exchange rate may be unappealing for an international audience. Nevertheless, starting with the influential paper of Chen, Rossi and Rogoff (2010), a growing literature has shown that the Chilean peso has the ability to Granger cause copper prices, base metal prices, and a World Commodity index. See for instance, Chen, Rossi and Rogoff (2010, 2014), Pincheira and Hardy (2018a) and the references cited therein. The potential finding of good predictors of the Chilean exchange rate may also illuminate the road to find good predictors for some of these commodity prices and, therefore, may result appealing for a worldwide audience[1].

[1] In fact, Pincheira and Hardy (2018b) show some interesting results of predictability from survey-based forecasts of the Chilean peso to base metal prices.

We are clearly not the first evaluating the predictive performance of survey-based forecasts of exchange rates. For instance, in the case of Mexico, Capistrán and López-Moctezuma (2010) show that, despite of being inefficient, survey-based forecasts outperform the DRW at several horizons in the last period of their sample, although mixed results are reported for the full sample. Ince and Molodtsova (2016) make an exhaustive analysis, considering 33 developed and developing countries, including Chile. Do they beat the random walk? Sometimes. Especially good results are found for developed countries at long horizons. In the particular case of Chile, results are not that impressive, as only one of the two surveys analyzed in that paper is able to outperform the DRW at one particular forecasting horizon: three months ahead. Ince and Molodtsova (2016) mention three additional articles exploring a similar subject, MacDonald and Marsh (1994, 1996) and Mitchell and Pearce (2007). The focus of these papers is on parities of a few advanced countries vis-a-vis the U.S. dollar. Generally speaking, in these papers the DRW is seldom outperformed.

Unlike Ince and Molodtsova (2016), we consider a different survey to obtain expectations of the Chilean exchange rate. We consider the SPF that has been conducted by the Central Bank of Chile since 2000 on a monthly basis. It is important to remark that in our sample period, Chile has had a floating exchange rate with only a handful of observations influenced by preannounced Central Bank interventions[2].

Our main results indicate that: 1. The SPF outperforms the DRW in terms of Mean Squared Prediction Error (MSPE) 2. This survey even outperforms a more competitive benchmark based on a refined information set. 3. The SPF also outperforms a "pure luck" forecast in terms of Directional Accuracy (DA).

Our findings, in combination with those of Capistrán and López-Moctezuma (2010) and Ince and Molodtsova (2016), suggest that survey-based forecasts of exchange rates should be

[2] See Pincheira (2018) for some description of such intervention periods.

considered as a tough benchmark to beat for economic models. In fact this benchmark is toughest than the traditional DRW in some countries like Chile.

The rest of this paper is organized as follows: Section 2 describes our data set. In Section 3 we evaluate the accuracy of survey-based forecasts in terms of MSPE. In Section 4 we focus on directional accuracy. Section 5 concludes.

2. Data

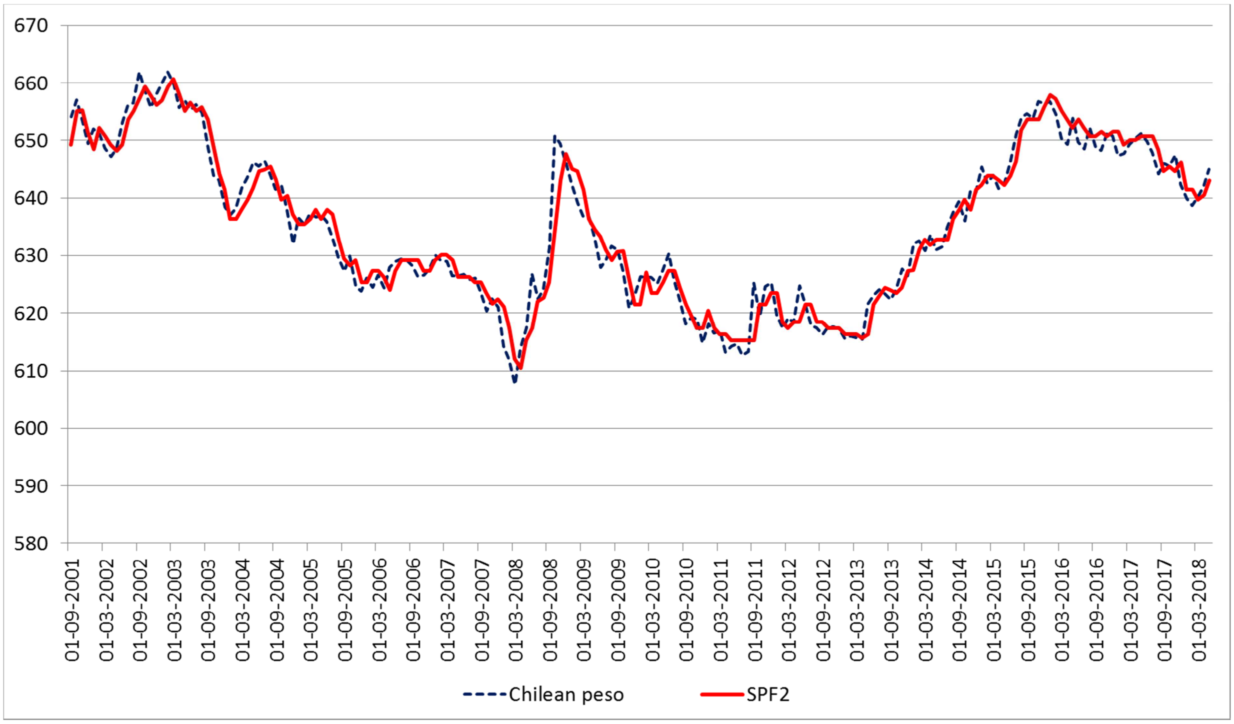
We use monthly data from September 2001 to May 2018. Our first source comes from the SPF released by the Central Bank of Chile. This survey targets scholars, consultants, and executives of the financial sector. Its purpose is to get to know the economic expectations of key economic variables like inflation, interest rates, exchange rates and GDP growth. The Central Bank of Chile releases the median values provided by the respondents. Pedersen (2010) gives a thorough description of the SPF.

The respondents of the survey are asked to predict the value of the exchange rate at three different horizons: 2, 11 and 23 months ahead (henceforth, SPF2, SPF11, SPF23). We go beyond these natural forecast horizons to evaluate the possibility that the time series on SPF2 and SPF11 may also be useful to predict exchange rates at a variety of forecasting horizons. This might sound counterintuitive, but this strategy is inspired on the fact that the optimal forecast of a DRW is the same for every single horizon. We notice here that our analysis focuses only on SPF2 and SPF11, leaving the analysis of SPF23 as a possible extension for future research[3].

We extract Chilean daily exchange rates from Bloomberg (last price). Our data are converted to monthly frequencies by sampling from the last day of the month. Figure 1 shows the SPF2 (survey of professional forecasters two months ahead) and the Chilean exchange rate. We clearly see how closely the survey tracks the Chilean peso.

[3] Our preliminary analysis, however, reveals that the predictive performance of SPF23 is not very promising.

Figure 1: Evolution of the American Dollar in Terms of Chilean Pesos and Survey-Based Forecast Two Months Ahead (SPF2)



Note: Data obtained from the Central Bank of Chile and Bloomberg. The dotted line represents the value of one American dollar in terms of Chilean pesos. For instance, at the beginning of our sample period, one dollar was equivalent to 654 Chilean pesos. The solid line represents the SPF forecast two months ahead.

3. Forecast Evaluation

In this section we compare the accuracy of the survey relative to the DRW in terms of MSPE. We also evaluate the stability of our results.

We compare the performance of the survey with forecasts coming from the simple DRW defined as:

$$s_{t+1} = s_t + e_{t+1}$$

where e_{t+1} is a white noise process and S_t is the nominal exchange rate at period t measured as the amount of Chilean pesos required to buy an American dollar in the domestic market. As usual, lower-case letters denote the natural logarithm of the respective variable:

$$s_t \equiv \ln(S_t)$$

Our target is the *h*-period return defined as follows:

$$r_{t,t+h} = s_{t+h} - s_t$$

In our notation, h denotes the relevant forecast horizon in months. We consider *h=1, 3, 6, 9, 11, 12, 18, 24.*

Assuming a DRW model, the optimal linear forecast for $r_{t,t+h}$ is exactly zero. Using the survey, the corresponding forecast for $r_{t,t+h}$ is denoted by $r_{t^+}^{SPF}(h)$ and defined by

$$r_{t^+}^{SPF}(h) = s_{t^+}^{SPF}(h) - s_t$$

where $s_{t^+}^{SPF}(h) \equiv \ln(S_{t^+}^{SPF})$ for all horizon h. Here $S_{t^+}^{SPF}$ represents the forecast of the nominal exchange rate coming from the survey. We use the subscript "t^+" to explicitly remark that the respondents of the survey are required to provide their forecasts approximately on the 10th day of month "t+1". Given that they have a few days of information from month "t+1" to build their forecasts, we consider inadequate the subscripts "t" or "t+1" for the survey, as S_t represents the nominal exchange rate corresponding to the last day of month "t" and S_{t+1} represents the nominal exchange rate corresponding to the last day of month "t+1".

More generally, and considering that the timeline in the flow of information is given by the following relationship:

$$t < t^+ < t + 1$$

We will make use of the subscript " t^+ "to define a forecast constructed with a refined set containing the information available up until the day before the survey is released. For example, if the SPF was released the 10th day of a given month, the information set I_{t^+} contains information available up until the 9th day of that month.

We recall here that we will work with two different versions of the survey that we label SPF2 and SPF11. We will evaluate the ability of each one of these two versions to forecast the nominal exchange rate at several forecasting horizons.

The forecast error when forecasting with the DRW is given by

$$e_t^{DRW}(h) = r_{t,t+h} - 0 = r_{t,t+h} = s_{t+h} - s_t$$

The forecast error when forecasting with the SPF is given by

$$e_{t^+}^{SPF}(h) = r_{t,t+h} - \left[s_{t^+}^{SPF}(h) - s_t\right] = s_{t+h} - s_{t^+}^{SPF}(h)$$

To evaluate forecast accuracy under quadratic loss, we focus on the difference

$$\Delta MSPE_h = E[e_t^{DRW}(h)]^2 - E\left[e_{t^+}^{SPF}(h)\right]^2 \quad (1)$$

The following null hypothesis:

$$H_0 \colon \Delta MSPE_h \leq 0$$

is evaluated against the alternative

$$H_1 \colon \Delta MSPE_h > 0$$

Our null hypothesis posits that the DRW is at least as accurate as the survey. The alternative hypothesis indicates that the survey outperforms the DRW in terms of MSPE.

We use a one-sided Diebold and Mariano (1995) and West (1996) test (henceforth, DMW) in the spirit of Giacomini and White (2006) to evaluate our hypothesis, using HAC standard errors according to Newey and West (1987, 1994).

In Table 1 we show the results of this predictive evaluation for SPF2 and SPF11. In particular, Table 1 shows the Root Mean Squared Prediction Error (RMSPE) ratios between survey-based forecasts and the DRW. If the ratio is lower than one, the survey is more accurate than the DRW. This table also shows the t-statistic and the p-value of the DMW test. A positive value of the t-statistic favors the survey.

Table 1 shows striking results as SPF2 outperforms the DRW at every horizon at tight significance levels. These results are stronger than those reported by Ince and Molodtsova (2016) for Chile. In contrast, we cannot reject the null of better performance of the DRW relative to SPF11 at any single horizon. Notice, however, that RMSPE ratios for SPF2 and SPF11 are similar at horizons longer than 3 months. The implication is that in the DMW test, the reason behind the no rejection of the null for the SPF11 relies on high standard errors.

Table 1: Forecast accuracy of survey-based forecasts relative to the DRW at several forecasting horizons.

	h = 1			h = 2			h = 3		
	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value
SPF2	0.901	3.02	0.001	0.924	3.16	0.001	0.948	2.76	0.003
SPF11	1.121	-1.58	0.942	1.009	-0.14	0.558	0.984	0.29	0.388

	h = 6			h = 9			h = 11		
	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value
SPF2	0.964	3.12	0.001	0.957	2.93	0.002	0.962	3.03	0.001
SPF11	0.962	0.73	0.231	0.943	1.07	0.142	0.955	0.97	0.165

	h = 12			h = 18			h = 24		
	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value
SPF2	0.957	3.65	0.000	0.973	2.84	0.002	0.975	3.10	0.001
SPF11	0.953	1.05	0.146	0.970	0.96	0.167	0.973	1.06	0.146

Note: Data obtained from the Central Bank of Chile and Bloomberg. The DMW test is constructed with HAC standard errors. RMSPEs lower than 1 favor survey-based forecasts.

It is important to remark here that results in Table 1 show that SPF2 consistently outperform the DRW across different forecasting horizons. This is relevant, because at least part of the literature consider that some forecasts have been successful in outperforming the DRW at long horizons, but not very much at short horizons. See for instance, Rogoff and Stavrakeva (2008), Ince and Molodtsova (2016) and the references cited therein. Our results differ from those papers showing that SPF2 is able to outperform the DRW both in the short and in the long run.

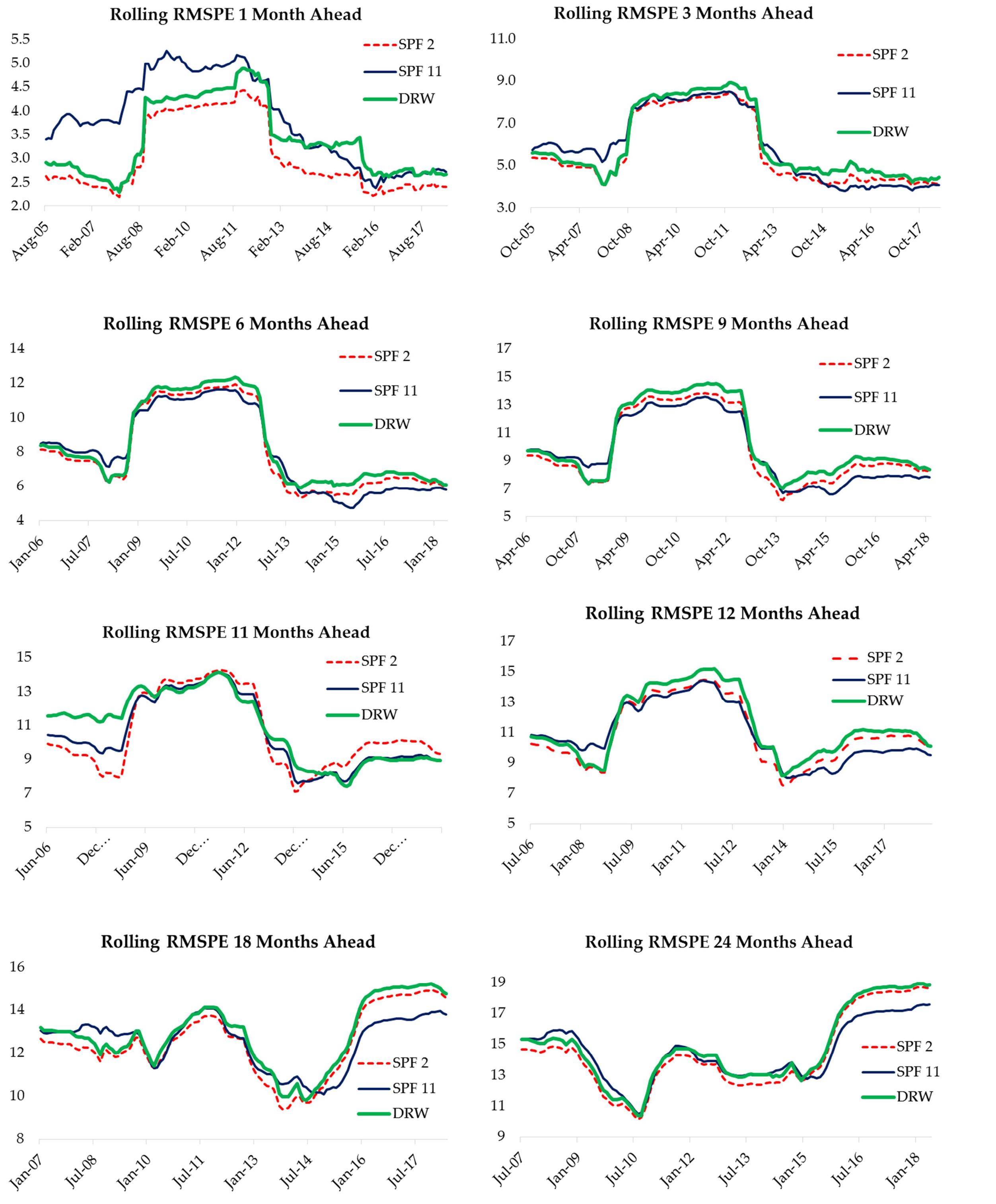
Another traditional result in the forecasting literature in general, and in the exchange rate literature in particular, has to do with instabilities. Rossi (2013b) presents a vast review documenting the existence of only sporadic episodes of predictability. Rossi (2013a) also remarks this feature in the particular case of the exchange rate forecasting literature. Similarly, Rogoff and Stavrakeva (2008) argue that their literature review reveals than often when a model outperforms the DRW it does so for a period of time, but in other subsamples, the opposite result emerges.

Accordingly, and to check for robustness, we explore the stability of our results visually in Figure 2[4]. This graph shows the RMSPE of the SPF2, SPF11 and the DRW in rolling windows of 48 observations.

Figure 2 is consistent with the literature in the sense that all three forecasts (SPF2, SPF11 and DRW) show time-varying forecast accuracy. Nevertheless, Figure 2 also shows that the RMSPE of the DRW is seldom the most accurate, as in most rolling windows, the RMSPE of the DRW is outperformed by both or at least one of the versions of the survey.

[4] A formal application of the DMW in a smaller subsample is introduced later in the paper.

Figure 2: **Rolling RMSPE of SPF2, SPF11 and the DRW at several horizons.**



Note: RMSPE calculated using rolling windows of 48 months. Due to the small scale of the errors, they are multiplied by 100.

One possible explanation for the outstanding predictive performance of SPF2 may be related to the richer information set available at the moment that the survey is conducted. Given that the respondents are required to provide their forecasts approximately on the 10th day of each month, the DRW is in clear information disadvantage relative to the survey. To explore this possible explanation, we consider a smaller sample for which we know the exact day in which the survey was released to the public. With this additional information, we construct a theoretically much more competitive benchmark based on the information set I_{t^+} defined previously. We denote this new benchmark by DRW^+. The more competitive forecast of $r_{t,t+h}$ is defined simply as

$$\mathrm{s}_{\mathrm{t}^+} - s_t$$

where $\mathrm{s}_{\mathrm{t}^+}$ represents the natural logarithm of the Chilean peso from the day before the SPF was released to the public.

The forecast error when forecasting with the DRW^+ is given by

$$e_{\mathrm{t}^+}^{DRW^+}(h) = r_{t,t+h} - (\mathrm{s}_{\mathrm{t}^+} - s_t) = s_{t+h} - \mathrm{s}_{\mathrm{t}^+}$$

To evaluate forecast accuracy under quadratic loss with this new benchmark, we focus on the difference

$$\Delta MSPE_h^+ = E\left[e_{\mathrm{t}^+}^{DRW^+}(h)\right]^2 - E\left[e_{t^+}^{SPF}(h)\right]^2$$

The null hypothesis:

$$H_0: \Delta MSPE_h^+ \leq 0$$

is evaluated against the alternative:

$$H_1: \Delta MSPE_h^+ > 0$$

using a one-sided DMW test as before[5].

Table 2 summarizes our analysis with this more competitive benchmark in the shorter sub-sample period for which we know the exact day in which the survey was released to the public. If we consider a significance level of 10%, SPF2 outperforms the DRW^{+}at 3 and 24 months only, whereas the SPF11 outperforms the benchmark at most horizons, except the first two. Notice also, that RMSPE ratios are lower for SPF11 than for SPF2 at horizons longer than 2 months.

At first glance Table 2 might seem at odds with Table 1 because, at some forecasting horizons, the performance of SPF2 and SPF11 relative to the corresponding benchmark gets reverted[6]. How can we understand these seemingly conflicting results? Figure 3 is part of the answer. In this figure, we depict the RMSPE of SPF2 and SPF11 in rolling windows of 48 months. This figure shows a time-varying relative behavior of both SPF2 and SPF11 at long horizons ($h > 2$). When we focus on horizons longer than 2 months, we see that in the first rolling windows SPF2 shows higher accuracy, but in the last rolling windows SPF11 performs better. This is consistent with the results in Table 2 that are built with the last 74 observations of our sample.

[5] We need to consider that the survey release date is available since April 2012. This means that in this analysis we only use the last 74 observations of our sample. Previous to April 2012 the exact day in which the survey was released is not available.

[6] In Table 1 the DMW test rejects the null in favor of SPF2 at every single horizon, but it never rejects in favor of SPF11. *Au contraire*, in Table 2, the DMW test rejects the null in favor of SPF2 in only two forecasting horizons, while rejecting in favor of SPF11 in almost every forecasting horizon.

Table 2: Forecast accuracy of survey-based forecasts relative to the DRW^{+} at several forecast horizons. Data since April 2012.

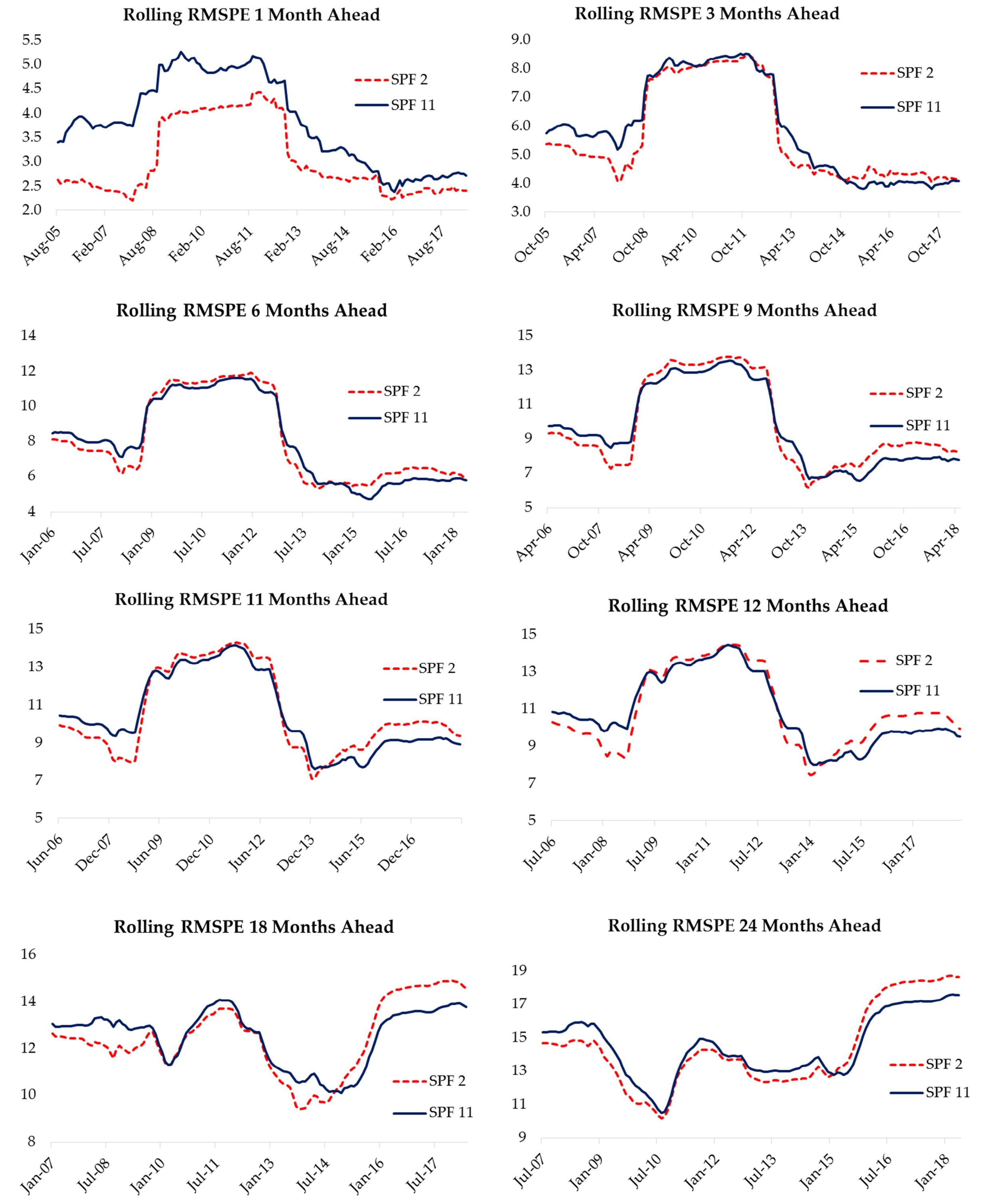
	h = 1			h = 2			h = 3		
	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value
SPF2	1.03	-0.57	0.715	0.97	0.84	0.200	0.96	1.31	0.095
SPF11	1.16	-1.68	0.953	0.97	0.39	0.347	0.92	1.30	0.096

	h = 6			h = 9			h = 11		
	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value
SPF2	0.97	1.28	0.101	0.98	0.93	0.176	0.98	0.87	0.193
SPF11	0.91	1.43	0.076	0.91	1.87	0.031	0.92	1.72	0.043

	h = 12			h = 18			h = 24		
	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value	RMSPE Ratio	t	p-value
SPF2	0.98	1.03	0.151	0.99	0.84	0.202	0.98	1.46	0.072
SPF11	0.92	1.69	0.045	0.93	1.94	0.026	0.93	2.83	0.002

Note: Data obtained from the Central Bank of Chile and Bloomberg. The DMW test is constructed with HAC standard errors. RMSPEs lower than 1 favor survey-based forecasts.

Figure 3: **Rolling RMSPE of SPF2 and SPF11 at several horizons.**



Note: RMSPE calculated using rolling windows of 48 months. Due to the small scale of the errors, they are multiplied by 100.

To complement our analysis, we show the RMSPE of SPF2, SPF11, DRW and DRW^{+} in both samples in Table 3. Due to the small scale of the errors, they are multiplied by 100.

Interestingly, Table 3 reveals that the DRW^{+} is indeed more accurate than the DRW at short and medium horizons. At long horizons their accuracy is almost the same. Table 3 also confirms that the accuracy of our forecasts is time-varying as the RMSPE is different in the whole sample and in the final sub-sample. In general our short-run forecasts are much more precise in the last subsample, but the long-run forecasts are substantially less precise in this last portion of our sample period.

Table 3: RMSPE at several horizons for the full sample and a more recent sub-sample.

	RMSPE (In Full Sample)			RMSPE (In Second Sub Sample)			
	SPF 2	**SPF 11**	**DRW**	**SPF 2**	**SPF 11**	**DRW**	**DRW +**
h = 1	**3.00**	**3.73**	**3.33**	**2.38**	**2.66**	**2.65**	**2.30**
h = 2	**4.49**	**4.90**	**4.86**	**3.29**	**3.31**	**3.64**	**3.40**
h = 3	**5.64**	**5.86**	**5.95**	**4.07**	**3.90**	**4.35**	**4.25**
h = 6	**8.02**	**8.01**	**8.32**	**5.86**	**5.53**	**6.05**	**6.04**
h = 9	**9.72**	**9.57**	**10.15**	**8.12**	**7.54**	**8.34**	**8.26**
h = 11	**10.38**	**10.30**	**10.64**	**9.35**	**8.73**	**9.54**	**9.49**
h = 12	**10.72**	**10.67**	**11.19**	**9.95**	**9.34**	**10.14**	**10.13**
h = 18	**12.42**	**12.38**	**12.76**	**14.10**	**13.23**	**14.26**	**14.29**
h = 24	**14.69**	**14.67**	**15.07**	**18.34**	**17.29**	**18.51**	**18.63**

Note: Data obtained from the Central Bank of Chile and Bloomberg. RMSPE is amplified by 100. DRW+ corresponds to the benchmark with a more refined information set.

In summary, Tables 1-3 consistently show a better performance of the SPF (either SFP2 or SPF11) relative to both the traditional DRW and the more competitive DRW^{+}. This means that the superior performance of the SPF cannot be justified merely in terms of an informational advantage relative to the DRW. While it is not the purpose of this paper to carry out an in-depth analysis of the causes of this superior performance, our results are consistent with the idea that the Chilean exchange rate is predictable.

4. Directional Accuracy

In this section we analyze the ability of the survey to forecast the future direction of the currency. Based on Moosa and Burns (2016) we use the following measure of Directional Accuracy:

$$DA_h^{SPF} \equiv \frac{1}{T}\sum_{t=1}^{T} z_t^{SPF}(h)$$

Where

$$z_t^{SPF}(h) = \begin{cases} 1 & if \quad (s_{t+h} - s_t)\big(s_{t^+}^{SPF}(h) - s_t\big) > 0 \\ 0 & \quad\;\; (s_{t+h} - s_t)\big(s_{t^+}^{SPF}(h) - s_t\big) \leq 0 \end{cases} \qquad (2)$$

And T represents the number of available observations.

To evaluate Directional Accuracy, we consider the following hypotheses:

$$H_0: E(z_t^{SPF}(h)) \leq \; 0.5$$

$$H_A: E(z_t^{SPF}(h)) > \; 0.5$$

The null hypothesis posits that the survey rate of success in predicting the future direction of the exchange rate is equal or lower than 0.5. In other words, the survey is unable to predict the future direction of exchange rates better than a "pure luck" mechanism[7]. Rejecting the null hypothesis means that the rate with which the survey correctly predicts the change of direction is greater than 50%. We test the null with a straightforward one-sided t-statistic using HAC standard errors[8].

[7] For instance, flipping a balanced coin.

[8] Cheung, Chinn, and Pascual (2005) consider a similar test.

Table 4 shows our results following the same structure of Table 1 for the full sample of observations. In particular, Table 4 exhibits the DA, also called “hit rate”, the t-statistic and the p-value for both SPF2 and SPF11 at several forecasting horizons. A positive value of the t-statistic means that the survey hit rate is greater than 50%.

Both versions of the survey display hit rates higher than 50% with only one exception (SPF11, *h* = 3). Moreover, in the case of SPF2, the hit rate is above 60% for all forecasting horizons. Accordingly, the null hypothesis of a “pure luck” mechanism is rejected for SPF2 at tight significance levels, with no exception. For the SPF11 the hit rate is, in general, not statistically significant. Only when forecasting 2 months ahead the null is rejected in favor of the SPF11 at the 10% significance level.

Table 4: Hit rate of survey-based-forecasts at several horizons during the full sample.

	h = 1			h = 2			h = 3		
	DA	t	p-value	DA	t	p-value	DA	t	p-value
SPF2	65%	5.62	0.000	68%	4.66	0.000	64%	3.56	0.000
SPF11	54%	1.05	0.148	56%	1.30	0.096	48%	-0.37	0.643
	h = 6			h = 9			h = 11		
	DA	t	p-value	DA	t	p-value	DA	t	p-value
SPF2	65%	3.85	0.000	65%	3.62	0.000	63%	2.81	0.002
SPF11	52%	0.35	0.362	58%	1.13	0.129	57%	1.02	0.154
	h = 12			h = 18			h = 24		
	DA	t	p-value	DA	t	p-value	DA	t	p-value
SPF2	63%	2.98	0.001	62%	2.40	0.008	64%	3.16	0.001
SPF11	56%	0.93	0.177	58%	1.09	0.137	55%	0.68	0.250

Note: Data obtained from the Central Bank of Chile and Bloomberg. We use HAC standard errors according to Newey and West (1987, 1994).

In line with Ince and Molodtsova (2016), we also evaluate the directional accuracy of survey forecasts with the nonparametric test proposed by Pesaran and Timmermann (1992). The null hypothesis of this test is that of independency between forecasts and the target variable. If they are independent, the forecast cannot have the ability to predict the sign of the predictand. We consider a one-sided version of this test, rejecting the null only when the hit rate is higher than the estimator of its expected value under the null. Table A in the appendix displays the results of this test. For SPF2 they are roughly consistent with those reported in Table 4, while for SPF11 they are much stronger, providing more evidence of sign predictability for the Chilean exchange rate.

The ability of SPF2 to detect the direction in which the exchange rate will move is outstanding relative to a "pure luck" benchmark (see Table 4). One possible explanation for this good performance may be related to the fact that the respondents of the survey already know a fraction of the predictand $r_{t,t+h}$. Let us recall that our target variable can be written as follows:

$$r_{t,t+h} \equiv s_{t+h} - s_t = [s_{t+h} - \mathrm{s}_{\mathrm{t}^+}] + [\mathrm{s}_{\mathrm{t}^+} - s_t]$$

Given that the respondents of the survey build their forecasts with information up until time "t^+", they know for sure the component $[\mathrm{s}_{\mathrm{t}^+} - s_t]$ of the target variable, which might be influential in the construction of the forecast of the future direction of the total return $r_{t,t+h}$.

To explore this possible explanation, we make use of the smaller sample for which the exact day in which the survey is released to the public is available (same subsample used in Table 2). We redefine our target variable as

$$r_{t+h}^{+} \equiv s_{t+h} - \mathrm{s}_{\mathrm{t}^+}$$

and define the following refined survey-based forecast for r_{t+h}^{+}:

$$r_{t^{++}}^{SPF}(h) = s_{t^+}^{SPF}(h) - \mathrm{s}_{\mathrm{t}^+}$$

When forecasting the change of direction of r_{t+h}^{+} with information based on the refined predictor $r_{t^{++}}^{SPF}(h)$, we consider the following measure of Directional Accuracy:

$$DA_h^{SPF^+} \equiv \frac{1}{P}\sum_{t=1}^{P} z_{t^+}^{SPF}(h)$$

where

$$z_{t^+}^{SPF}(h) = \begin{cases} 1 & if \quad (s_{t+h} - s_t^+)\left(s_{t^+}^{SPF}(h) - s_t^+\right) > 0 \\ 0 & \quad\quad (s_{t+h} - s_t^+)\left(s_{t^+}^{SPF}(h) - s_t^+\right) < 0 \end{cases}$$

and P represents the number of available observations.

Table 5 shows very interesting results. While the hit rate of SPF2 is higher than 50% in every horizon, it is statistically significantly better than the benchmark only at the short horizons of 1 and 2 months. Results for SPF11 are stronger. Its hit rate is always higher than 57% and the null of equal directional accuracy is rejected in favor of SPF11 at usual significance levels with only two exceptions.

Relative to Table 4, results in Table 5 are weaker for SPF2 but stronger for SPF11. More importantly, we see that the refinement in the target variable and in the forecast does not destroy the predictability of the survey. While results in Tables 4 and 5 are not directly comparable due to the use of different samples, they document a relevant ability of the SPF (either SFP2 or SPF11) to predict the future direction of the Chilean exchange rate.

Table 5: Hit rate of survey-based-forecasts at several horizons using a shorter subsample and a refined target variable and forecast.

	h = 1			h = 2			h = 3		
	DA	t	p-value	DA	t	p-value	DA	t	p-value
SPF2	57%	1.40	0.081	59%	1.80	0.036	56%	1.21	0.113
SPF11	62%	2.26	0.012	64%	2.51	0.006	67%	3.16	0.001

	h = 6			h = 9			h = 11		
	DA	t	p-value	DA	t	p-value	DA	t	p-value
SPF2	57%	1.04	0.150	55%	0.62	0.269	59%	1.14	0.127
SPF11	58%	1.03	0.151	62%	1.43	0.076	67%	1.81	0.035

	h = 12			h = 18			h = 24		
	DA	t	p-value	DA	t	p-value	DA	t	p-value
SPF2	57%	0.77	0.220	53%	0.27	0.393	57%	0.72	0.237
SPF11	63%	1.28	0.100	68%	1.70	0.045	76%	3.37	0.000

Note: Data obtained from the Central Bank of Chile and Bloomberg. We use HAC standard errors according to Newey and West (1987, 1994).

5. Concluding Remarks

Using a monthly database, in this paper we show that survey-based forecasts of the Chilean exchange rate vis-a-vis the US dollar consistently outperform the Driftless Random Walk (DRW) in terms of Mean Squared Prediction Error at several forecasting horizons, including both the short and long-run. Following Goyal and Welch (2008), our equivalent measures of out-of-sample goodness of fit are in the range of 5%-19%. We also show that survey-based forecasts are able to outperform an even more competitive benchmark than the DRW. This benchmark is constructed using a refined information set based on daily data. We report similar results when precision is measured in terms of Directional Accuracy: our survey-based forecasts outperform a "pure luck" benchmark at several forecasting horizons.

To our surprise, and differing from the traditional "no predictability" result found in the literature for many exchange rates, our striking findings clearly support the hypothesis that the Chilean peso is indeed predictable.

It is important to emphasize here that our analysis is based on a straightforward out-of-sample methodology: we compare the no change forecast resulting from the Driftless Random Walk model to real-time exchange rate forecasts coming from the Survey of Professional Forecasters carried out by the Central Bank of Chile on a monthly basis. In our exercise there is no need for parameter estimation. Inference is carried out using the traditional Diebold and Mariano (1995) and West (1996) test. Our total number of observations is 201. Furthermore, during our sample period, Chile had a free float with only a handful of preannounced intervention periods carried out by the monetary authority. We also explore the stability of our results analyzing a shorter subsample with the last 74 observations of our sample period. In sum, our findings are neither the result of a novel econometric artifact nor the results of a magic black box. They are plain and strong, yet surprising giving the long tradition of frustration with economics models, and some surveys too, when it comes to compare their forecasts with those of the simple random walk.

It is probably true that the vast majority of the research in exchange rate forecasting, including the seminal contributions of Meese and Rogoff (1983a, 1983b), focus on advanced economies, yet to our knowledge there is no particular economic reason to expect a different outcome in a country like Chile, that has followed an inflation targeting regime with a free float during all our sample period. Furthermore, Chen, Rossi and Rogoff (2010) report that, for the specific case of Chile and other commodity exporter countries, the Meese-Rogoff puzzle also holds true. They consider as potential predictors standard exchange rate fundamentals plus lags of the returns of a country-specific commodity or commodity index.

The natural question to ask here is: what is driving the predictability of the survey? On the one hand we have analyzed the median of the respondents, which is a particular type of forecast combination. It might be the case that at least some part of the results we have reported may be associated to the aggregation of a number of educated forecasts. Unfortunately, individual responses are not publicly available, so it is difficult to empirically explore this hypothesis. On the other hand, it might be the case that at least a few of the respondents of the survey may be basing their forecasts on a particular collection and combination of fundamentals. The identification of that collection and particular combination seems a fruitful avenue for future research.

References

1. Capistrán, C. and G. López-Moctezuma (2010): "Las Expectativas Macroeconómicas de los Especialistas: Una Evaluación de Pronósticos de Corto Plazo en México". *El trimestre económico* 77 (2): 275 – 312. (In Spanish)

2. Chen Y., K. Rogoff and B. Rossi (2010): "Can Exchange Rates Forecast Commodity Prices?". *The Quarterly Journal of Economics* 125 (3), 1145 – 1194.

3. Chen Y., K. Rogoff and B. Rossi (2014): Can Exchange Rates Forecast Commodity Prices? An Update, manuscript, February 2014.

4. Cheung, Yin-Wong, M. D. Chinn and A. G. Pascual (2005): "Empirical exchange rate models of the nineties: Are any fit to survive?" *Journal of International Money and Finance* 24: 1150 – 1175.

5. Clark, T. E. and K.D. West (2006): "Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis". *Journal of Econometrics* 135: 155 – 186.

6. Diebold, F. X. and R. S. Mariano (1995): "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13 (3): 134 – 144.

7. Engel, C. and K. D. West (2005): "Exchange Rates and Fundamentals". *Journal of Political Economy* 113 (3): 485 – 517.

8. Engel, C, N. Mark and K.D. West (2015): "Factor Model Forecasts of Exchange Rates". *Econometric Reviews*, 34(1–2):32–55, 2015

9. Giacomini R. and H. White (2006): "Tests of Conditional Predictive Ability," *Econometrica,* Econometric Society, vol. 74(6), pages 1545-1578, November.

10. Goyal A. and I. Welch (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455--1508.

11. Ince, O. and T. Molodtsova (2016): "Rationality and forecasting accuracy of exchange rate expectations: Evidence from survey-based forecasts". *Journal of International Financial Markets, Institutions & Money 47*: 131 – 151.

12. MacDonald, R. and I. Marsh (1994). "Combining exchange rate forecasts: what is the optimal consensus measure?" *International Journal of Forecasting*. 13, 313–332.

13. MacDonald, R. and I. Marsh (1996). "Currency forecasters are heterogeneous: confirmation and consequences". *Journal of International Money and Finance.* 15, 665–685.

14. Meese, R. A. and K. Rogoff (1983a): "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?". *Journal of International Economics* 14: 3 – 24.

15. Meese, R. A. and K. Rogoff (1983b): "The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?". *Journal of International Economics* 14: 3 – 24.

16. Mitchell, K. and D. Pearce (2007). "Professional forecasts of interest rates and exchange rates: evidence from the Wall Street Journal's panel of economists". *Journal of Macroeconomics*. 29, 840–854.

17. Molodtsova, T. and D. Papell (2009). "Out-of-sample exchange rate predictability with Taylor rule fundamentals," *Journal of International Economics*, Elsevier, vol. 77(2), pages 167-180, April.

18. Moosa, I. and K. Burns (2016): "The random walk as a forecasting benchmark: drift or no drift?" *Applied Economics*. Volume 48, 2016 - Issue 43. Pages 4131-4142

19. Newey, W.K. and K.D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". *Econometrica* 55 (3): 703-708.

20. Newey, W.K. and K.D. West (1994): "Automatic lag selection in covariance matrix estimation". *Review of Economic Studies* 61 (4): 631-654.

21. Pedersen, M. (2010): "Una nota introductoria a la encuesta de Expectativas Económicas". *Estudios Económicos Estadísticos* Nº 82, Banco Central de Chile. (In spanish).

22. Pesaran, M.H. and. A. Timmermann (1992). "A Simple Nonparametric Test of Predictive Performance" *Journal of Business & Economic Statistics* 10 (4): 461 – 465.

23. Pincheira, P. (2018). "Interventions and Inflation Expectations in an Inflation Targeting Economy," *Economic Analysis Review*, forthcoming.

24. Pincheira, P. and N. Hardy (2018a). "Forecasting Base Metal Prices with Commodity Currencies," MPRA Paper 83564, University Library of Munich, Germany.

25. Pincheira, P. and N. Hardy (2018b). "The Predictive Relationship Between Exchange Rate Expectations and Base Metal Prices". MPRA Paper 89423, University Library of Munich, Germany.

26. Rossi, B. (2013a): "Exchange Rate Predictability". *Journal of Economic Literature* 51 (4): 1063 – 1119.

27. Rossi, B. (2013b): "Advances in Forecasting under Instability". Handbook of Economic Forecasting Volume 2, Part B. Pages 1203-1324.

28. Rogoff K. and V. Stavrakeva (2008). "The Continuing Puzzle of Short Horizon Exchange Rate Forecasting," NBER Working Papers 14071, National Bureau of Economic Research, Inc.

29. West K. D. (1996): "Asymptotic Inference about Predictive Ability." *Econometrica* 64 (5): 1067 – 1084.

Appendix: Results of the nonparametric test of predictive performance

Like Ince and Molodtsova (2016), we also evaluate the directional accuracy of the survey with a nonparametric test based on Pesaran and Timmermann (1992). The null hypothesis of this test is that of independency between the forecast and the predictand. If they are independent, the forecast cannot have the ability to predict the sign of the target variable. The test is based on the following $S_T(h)$ statistic:

$$S_T(h) = \frac{\hat{p}_h - \hat{p}_h^*}{\left(\hat{v}(\hat{p}_h) - \hat{v}(\hat{p}_h^*)\right)^{1/2}}$$

Where T is the available sample size and $\hat{p}_h$ is the hit rate (proportion of times that the sign of $r_{t,t+h}$ is correctly forecasted). In addition, $\hat{p}_h^*$ is an estimator of the expectation of $\hat{p}_h$ under the null, computed according to the following expression:

$$\hat{p}_h^* = \hat{p}_{y(h)}\hat{p}_{x(h)} + (1 - \hat{p}_{y(h)})(1 - \hat{p}_{x(h)})$$

where $\hat{p}_{y(h)}$represents the proportion of times in which the target variable $r_{t,t+h}$ is positive in our sample period, and similarly, $\hat{p}_{x(h)}$ represents the proportion of times in which the forecast $s_{t^+}^{SPF}(h) - s_t$ is positive in our sample period. Pesaran and Timmermann (1992) also provide expressions for the variance terms:

$$\hat{v}(\hat{p}) = \frac{1}{T}\hat{p}_h^*(1 - \hat{p}_h^*)$$

$$\hat{v}(\widehat{p_*}) = \frac{1}{T}\left(2\hat{p}_{y(h)} - 1\right)^2 \hat{p}_{x(h)}\left(1 - \hat{p}_{x(h)}\right) + \frac{1}{T}\left(2\hat{p}_{x(h)} - 1\right)^2 \hat{p}_{y(h)}\left(1 - \hat{p}_{y(h)}\right) + \\ + \frac{4}{T^2}\hat{p}_{y(h)}\hat{p}_{x(h)}(1 - \hat{p}_{y(h)})(1 - \hat{p}_{x(h)})$$

Table A shows the results of a one sided version of the test. This means that we reject only when the hit rate is greater than $\hat{p}_h^*$. As we see, we strongly reject the null hypothesis at several forecasting horizons for both versions of the survey, SPF2 and SPF11. The only exception occurs when forecasting with SPF11 three months ahead. In this particular case, the null hypothesis cannot be rejected at usual significance levels.

Table A: Pesaran and Timmermann (1992) test at several horizons.

	h = 1			h = 2			h = 3		
	Hit Rate	$S_T(1)$	p-value	Hit Rate	$S_T(2)$	p-value	Hit Rate	$S_T(3)$	p-value
SPF2	65%	4.33	0.000	68%	5.11	0.000	64%	3.94	0.000
SPF11	54%	2.04	0.021	56%	2.49	0.006	48%	0.83	0.204

	h = 6			h = 9			h = 11		
	Hit Rate	$S_T(6)$	p-value	Hit Rate	$S_T(9)$	p-value	Hit Rate	$S_T(11)$	p-value
SPF2	65%	4.33	0.000	65%	4.30	0.000	63%	3.58	0.000
SPF11	52%	2.17	0.015	58%	3.86	0.000	57%	3.78	0.000

	h = 12			h = 18			h = 24		
	Hit Rate	$S_T(12)$	p-value	Hit Rate	$S_T(18)$	p-value	Hit Rate	$S_T(24)$	p-value
SPF2	63%	3.64	0.000	62%	3.22	0.001	64%	3.69	0.000
SPF11	56%	3.25	0.001	58%	4.05	0.000	55%	2.94	0.002

Note: Data obtained from the Central Bank of Chile and Bloomberg. $S_T(h)$ represents the statistic proposed by Pesaran and Timmermann (1992).