Predictability v/s Accuracy: An empirical evaluation of the Chilean Exchange Rate with the Chilean Survey of Professional Forecasters

TESIS PARA OPTAR AL GRADO DE MAGISTER EN ECONOMIA

30/09/2024

Martín Ignacio Flores Donoso

PROFESOR GUIA: Pablo Pincheira Brown

PROFESORES CORRECTORES: Francisco Parro y Lorenzo Reus

ACCREDITATIONS MEMBER OF

Predictability and bias adjustment of the Chilean Survey of Professional Forecasters for the Chilean Exchange Rate

Abstract

We evaluate the predictability and efficiency of the Survey of Professional Forecasters (SPF) of the Chilean exchange rate relative to the US dollar (CLP). We evaluate the forecast performance through four approaches: (1) a comparison with a naïve benchmark under using a quadratic loss function, (2) Mean Directional Accuracy relative to a pure luck benchmark, (3) the correlation between SPF and the target variable, and (4) a range analysis of the respondents. Our analysis spans from April 2012 to April 2024, with a specific focus on the period from June 2018 to April 2024. Out-of-sample results reveals that SPF is consistently outperformed by the Driftless Random Walk (DRW) in terms of Mean Squared Prediction Error at several forecasting horizons. Nevertheless, SPF correlations are strong and statistically significant at short, mid, and long horizons. Therefore, CLP is predictable but inaccurate, with the bias playing an important role. To address this, we propose a bias adjusted forecast. This adjusted forecast outperforms the more competitive benchmark, the DRW+, at mid and long horizons.

1. Introduction

In this paper we re-evaluate the ability of the SPF to predict the CLP exchange rate returns at several forecasting horizons. As demonstrated by Meese and Rogoff (1983), the DRW has proven to be a very difficult benchmark to outperform in out-of-sample evaluations within the exchange rate literature. Since then, extensive research has been devoted to overcoming this challenge. See for instance, Ince and Molodtsova (2017), Pincheira and Neumann (2022), Ren et al (2018) have achieve success in outperforming this benchmark. Besides that, Berg and Mark (2015) have managed to explain why macroeconomics fundamentals are not good to predict exchange rates. However, such findings remain exceptions rather than the norm. According to Rossi's (2013) comprehensive survey, these advancements have not been sufficient to overturn the conclusions of Meese and Rogoff. Generally, the DRW continues to be a formidable benchmark to beat.

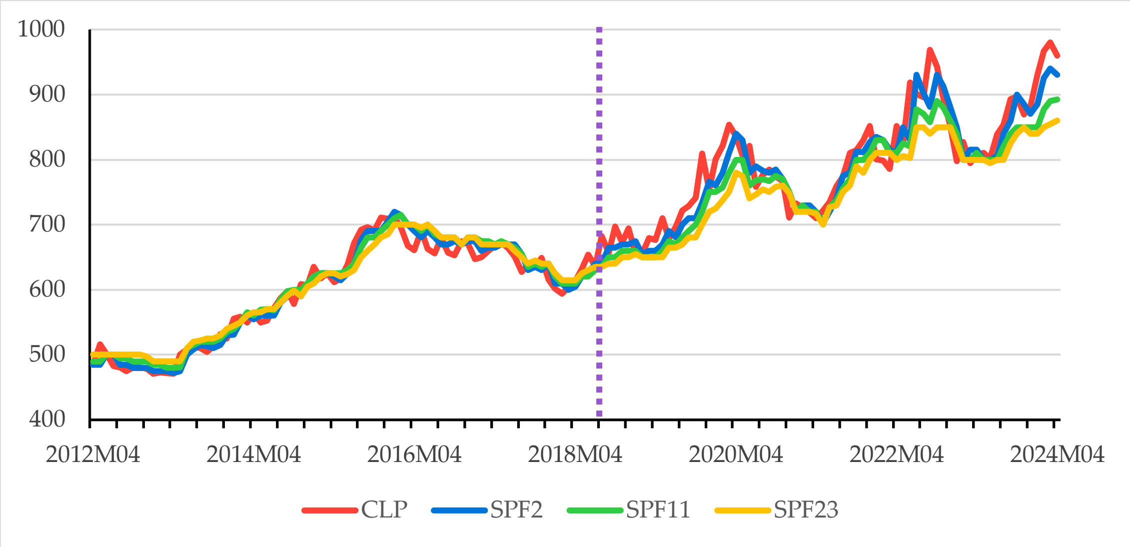
We also assess the efficiency of the survey, which is tested through the Mincer and Zarnowitz (1969) efficiency conditions. A rational forecast should be unbiased and linearly independent of its forecast error. Evaluating the rationality of survey-based forecasts is a common practice in the literature. On one hand, studies like Dominguez (1986), MacDonald and Marsh (1994) find that survey forecast are not efficient and lack predictive power, also Jongen et al. (2008) in their review of foreign exchange rate expectations report that are not rational and have low predictability. On the other hand, Capistrán and Moctezuma (2010) show that while the SPF conducted by the Central Bank of Mexico has predictability, is also inefficient and exhibit bias. Ince and Molodtsova (2017) test these conditions for 33 different countries. They also find that both Consensus and F4xcast exchange rate surveys are bias and inefficient, specially at forecasting horizons of 12 and 24 months.

This analysis is not entirely novel. Pincheira and Neumann (2022) shows that the SPF does, in fact, predict the CLP, outperforming the DRW at several horizons. However, we believe that is crucial to evaluate the performance of the SPF given the significant global and local exogenous events that have occurred, such as COVID-19, the Russia-Ukraine war, the 2019 Social Outbreak in Chile, and two national referendums. These events likely had a substantial impact on the CLP exchange rate, as illustrated in Figure 1, where the purple 90° line stands on June 2018. It is evident that both the CLP and SPF series have a marked change. For this reason, our analysis focuses on the period from June 2018 to April 2024.

Also, we believe that the benchmark that has been widely used in the survey-based forecast is not appropriate. Because survey-based forecasts have a greater information set than the DRW. In consequence, we propose a new benchmark which has a similar information set, the DRW+. Besides that, according to Pincheira and Hardy (2024), the predictability comes from correlation value, and not from accuracy measures such as MSPE. A feature that these past papers do not evaluate at all, focusing just on the precision instead.

Our main findings are: 1. The SPF is outperformed by both the DRW and DRW+ in terms of MSPE at several horizons 2. The direction of change accuracy is significantly reduced, with a hit rate below 50% at multiple horizons 3. Despite these, the SPF predictability remains intact. Presenting high and statistically significant correlations with the target variable at short, medium, and long forecasting horizons 4. SPF forecast bias has increase in the June 2018 to April 2024 period, at all horizons. Presenting a systematic underestimation of actual returns 5. We propose a bias adjusted (BA) forecast, corrected by bias. The BA forecast not only outperforms the more competitive benchmark at mid and long-term horizons but also dramatically improves direction-of-change accuracy.

Figure 1: Chilean exchange rate relative to US dollar and SPF forecasts in level. 2012M04 – 2024M04



Notes: The dotted line represents the division of our sample. 2018M06 on ward.

The rest of this paper is organized as follows: Section 2 describes our data set. Section 3 evaluates the survey accuracy and dependence by four different approaches. In Section 4 we analyze forecast efficiency and bias condition. Section 5 presents our new adjusted forecast by bias. We conclude in Section 6.

2. Data

We use monthly data from April 2012 to April 2024. As this is the only period where the Central Bank of Chile (CBCH) publishes the release dates of the survey.

Our first source comes from the SPF released by the CBCH. This survey is conducted during the first week of each month and released in the second week. It targets scholars, consultants, and executive of the financial sector. The CBCH releases median values and the 10th and 90th percentiles of the survey.

The survey asks for exchange rate forecasts at three different horizons: 2, 11 and 23 months ahead (SPF2, SPF11 and SPF23 from now on). Read Pedersen (2010) for further details on how the survey is constructed.

We extract the closing price of Chilean exchange rate (CLP) from Bloomberg. Our data are converted into monthly frequencies by sampling from the last day of the month. We also sample closing price of the day before survey release, which we call CLP+. Is plus because is in the middle of the month, then t<t+<t+1.

We take three evaluation windows, 2012M04 – 2024M04 (full sample), 2012M04 – 2018M05 (first subsample) and 2018M06 – 2024M04 (second subsample).

3. Forecast Evaluation

In this section, we evaluate the SPF performance using four distinct approaches: (1) forecast accuracy measured by Mean Squared Prediction Error (MSPE) relative to two benchmarks—the common Driftless Random Walk (DRW) and the more competitive DRW+, which shares a similar information set with the survey; (2) Mean Directional Accuracy compared to a pure luck benchmark; and (3) the correlation between the SPF and the target variable. And finally (4) the range coverage between the 10th and 90th percentiles.

These allows us to test SPF capabilities under accuracy measurements, MSPE, Sign and Range. And under dependence measurements, Sign and Correlations. Leading into a more robust evaluation of the forecast.

3.1 MSPE Forecast Evaluation

We compare SPF accuracy under MSPE relative to the DRW at several h forecasting horizons. With h = 1, 2, 3, 6, 9, 11, 12, 18 and 24 months ahead. Therefore, our forecast and our benchmark are the equation (1) and (2) respectively:

$$r_{t+}^{SPF}(h) = s_{t+}^{SPF}(h) - s_t \qquad (1)$$
$$r_t^{DRW} = 0 \qquad (2)$$

Where $s_t \equiv ln(S_t)$ and S_t is the spot exchange rate at time t. While $s_{t+}^{SPF} \equiv ln\left(S_{t+}^{SPF}\right)$ and S_{t+}^{SPF} are the forecast of nominal exchange rate at time t+. As in Pincheira and Neumann (2022) we use subscript t+ because survey results are published in the second week of each month, therefore t < t+ < t+1.

We evaluate if these forecasts are statistically better than the DRW with a Diebold and Mariano (1995) and West (1996) (DMW) test with HAC standard errors from Newey and West (1987).

When forecasting with the SPF and the DRW, our forecast errors are:

$$e_{t+}^{SPF}(h) = s_{t+h} - s_{t+}^{SPF} \qquad (3)$$
$$e_t^{DRW}(h) = s_{t+h} - s_t \qquad (4)$$

To evaluate forecast accuracy under MSPE, we focus on the difference $\Delta MSPE_h = E[e_t^{DRW}(h)]^2 - E\left[e_{t+}^{SPF}(h)\right]^2$. With the following hypotheses:

$$H_0: \Delta MSPE_h \leq 0$$
$$\mathrm{H_A}: \Delta\mathrm{MPSE_h} > 0$$

A null rejection implies that SPF forecast outperform the DRW at a statistically significant level. For inference, we apply a one-sided Diebold and Mariano (1995) and West (1996) test (referred to as the DMW test) using HAC standard errors according to Newey and West (1987, 1994). The results for the full sample set are presented in Table 1. While the results for the first and second subsample are shown in Table 2 and 3, respectively. The entries show Root Mean Squared Prediction Error (RMSPE) ratios between SPF and the DRW. Ratios below 1 favor survey-based forecast. The tables also display t-statistic and P-Value of the DMW test. These results are an out-of-sample evaluation.

Table 1 results show that we can reject 4 out of 27 cases. With SPF2 being the only survey that outperforms DRW at a statistically significant level for 1, 2, 3 and 6 months ahead.

Table 1: Forecast accuracy of survey-based forecasts relative to the DRW at several forecasting horizons. 2012M04 – 2024M04 window.

	h = 1			**h = 2**			**h = 3**		
	RMSPE Ratio	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**
SPF2	**0.9046*****	3.4592	0.004	**0.9367****	1.8962	0.0300	**0.9327****	1.8923	0.0302
SPF11	1.1562	-1.7052	0.0452	1.0953	-1.1630	0.1234	1.0184	-0.2644	0.3959
SPF23	1.4578	-2.5882	0.0053	1.2951	-2.1065	0.0185	1.1381	-1.2884	0.0998
	h = 6			**h = 9**			**h = 11**		
	RMSPE Ratio	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**
SPF2	**0.9655***	1.4580	0.0735	**0.9800***	1.3414	0.0910	**0.9920**	0.4908	0.3121
SPF11	**0.9777**	0.3652	0.3581	**0.9656**	0.7308	0.2330	**0.9819**	0.3966	0.3461
SPF23	1.0257	-0.3005	0.3821	**0.9863**	0.2139	0.4154	**0.9980**	0.0338	0.4865
	h = 12			**h = 18**			**h = 24**		
	RMSPE Ratio	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**
SPF2	**0.9939**	0.4141	0.3397	1.001	-0.0862	0.4657	**0.9968**	0.4044	0.3433
SPF11	**0.9856**	0.3454	0.3651	**0.9957**	0.1260	0.4499	**0.9853**	0.5770	0.2824
SPF23	**0.9978**	0.0396	0.4842	1.0072	-0.1563	0.4380	**0.9908**	0.2245	0.4113

Notes: DMW test is constructed with HAC standard errors. RMSPEs lower than 1 favor survey-based forecasts. * Significance at 10%, ** significance at 5%, ***significance at 1%.

If we compare it to the first subsample results in Table 2, we find a totally different history. We reject null for 17 out of 27 entries at several horizons. SPF2 outperforms DRW at short-term, while SPF11 and SPF23 does the same at mid and long-term horizons.

The results in Table 3 further favor the DRW. We reject the null for just 2 entries, SPF2 at 1 and 3 months ahead. This accuracy drop could be attributed to the exogenous events that we mention in previous section and/or a structural change in the SPF. Regardless, the SPF accuracy has decreased when we compare it to the first subsample.

Table 2: Forecast accuracy of survey-based forecasts relative to the DRW at several forecasting horizons. 2012M04 – 2018M05 window.

	h = 1			h = 2			h = 3		
	RMSPE Ratio	T-stat	P-Value	RMSPE Ratio	T-stat	P-Value	RMSPE Ratio	T-stat	P-Value
SPF2	**0.8980******	2.3840	0.0098	**0.8943******	3.1073	0.00135	**0.9334****	1.7188	0.0450
SPF11	**0.9976**	0.0301	0.4880	**0.9105***	1.5439	0.0635	**0.8997***	1.5033	0.0685
SPF23	1.1467	-1.2297	0.1113	**0.9951**	0.0548	0.4782	**0.9609**	0.4557	0.3250
	h = 6			**h = 9**			**h = 11**		
	RMSPE Ratio	T-stat	P-Value	RMSPE Ratio	T-stat	P-Value	RMSPE Ratio	T-stat	P-Value
SPF2	**0.9687**	1.1568	0.1257	**0.9741***	1.5259	0.0659	**0.9806**	1.0511	0.1486
SPF11	**0.9135***	1.4203	0.0800	**0.9031****	2.1727	0.01672	**0.9148****	1.8231	0.0365
SPF23	**0.9259**	1.0899	0.1397	**0.8881****	2.1595	0.0172	**0.8955****	1.9149	0.0300
	h = 12			**h = 18**			**h = 24**		
	RMSPE Ratio	T-stat	P-Value	RMSPE Ratio	T-stat	P-Value	RMSPE Ratio	T-stat	P-Value
SPF2	**0.9820**	1.0013	0.16029	**0.9884**	0.8979	0.1915	**0.9905**	1.2132	0.1153
SPF11	**0.9205****	1.7355	0.0438	**0.9276****	2.1040	0.0199	**0.9341*****	2.9672	0.0023
SPF23	**0.9001****	1.8392	0.0353	**0.9022****	2.1957	0.0161	**0.9015*****	2.9864	0.0021

Notes: DMW test is constructed with HAC standard errors. RMSPEs lower than 1 favor survey-based forecasts. * Significance at 10%, ** significance at 5%, ***significance at 1%.

Although the literature has recognized the DRW as one of the toughest benchmarks to outperform, this does not hold true for expectations surveys. Because forecasters have a richer information set than the DRW, due to having around 9 to 12 days more to answer the survey. In consequence, we may be overstating empirical evidence.

Therefore, we propose a new tougher benchmark than DRW, the DRW+[1]. Where forecast (5) and its forecast error (6) are estimated as follows:

$$r_{t^+}^{DRW+}(h) = s_{t+} - s_t \qquad (5)$$
$$e_{t^+}^{DRW+}(h) = s_{t+h} - s_{t^+} \qquad (6)$$

[1] We also test MSPE accuracy relative to the Random Walk with Drift Plus. Results are in Appendix A.

This new benchmark is constructed by taking the closing price of CLP from the day before the survey is released ($s_{t}+$). The logic is that, if you are an investor who reads the survey, you will create your expectations based on the last piece of information you have available. In this case, the closing price of the day before. In this way both will have similar information set, leading into a fair contest.

Table 3: Forecast accuracy of survey-based forecasts relative to the DRW at several forecasting horizons 2018M06 – 2024M04 window.

	h = 1			h = 2			h = 3		
	RMSPE Ratio	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**
SPF2	**0.9074*****	3.9644	0.0000	**0.9580**	0.8924	0.1875	**0.9324***	1.4912	0.0701
SPF11	1.2184	-1.9299	0.0288	1.1800	-1.7151	0.0453	1.0671	-0.7371	0.2317
SPF23	1.5734	-2.8656	0.0027	1.4264	-2.5396	0.0066	1.2087	-1.5371	0.0643
	h = 6			**h = 9**			**h = 11**		
	RMSPE Ratio	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**
SPF2	1.0703	-1.4508	0.0756	**0.9830**	0.8257	0.2058	**0.9992**	0.0397	0.4821
SPF11	1.1123	-1.2301	0.1113	**09973**	0.0441	0.4824	1.0221	-0.3700	0.3562
SPF23	1.1795	-1.4072	0.0818	1.0341	-0.3989	0.3455	1.0578	-0.7221	0.2363
	h = 12			**h = 18**			**h = 24**		
	RMSPE Ratio	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**
SPF2	1.0017	-0.0880	0.4650	1.0144	-0.8490	0.19936	1.0050	-0.3417	0.3667
SPF11	1.0268	-0.5044	0.3077	1.0639	-1.6144	0.0554	1.0490	-1.9321	0.0286
SPF23	1.0581	-0.8195	0.2076	1.1087	-2.0912	0.0200	1.0973	-2.3001	0.0121

Notes: DMW test is constructed with HAC standard errors. RMSPEs lower than 1 favor survey-based forecasts. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead.

We estimate ratios against this new benchmark for full sample period and both subsamples. Results are display in Table 4, 5 and 6.

As we expect from a more competitive benchmark, in Table 4 there are only 8 entries below 1 and no null rejection. While Table 5 displays quite the opposite. We found that survey-based forecast outperforms DRW+. We reject null for 13 out of 27 cases, highlighting SPF11 and SPF23 at long term horizons. Table 6 goes the same way as before. None of the entries are below 1 at any horizon whatsoever.

Table 4: Forecast accuracy of survey-based forecasts relative to the DRW+ at several forecasting horizons. 2012M04 – 2024M04 window.

	h = 1			h = 2			h = 3		
	RMSPE Ratio	**t**	**P-Value**	**RMSPE Ratio**	**t**	**P-Value**	**RMSPE Ratio**	**t**	**P-Value**
SPF2	1.0818	-1.3034	0.0972	1.0655	-1.5500	0.0616	1.0442	-1.4961	0.0684
SPF11	1.3828	-2.7145	0.0037	1.2459	-2.3427	0.0102	1.1402	-1.6682	0.0487
SPF23	1.7434	-3.0114	0.0015	1.4731	-2.6689	0.0042	1.2741	-2.0697	0.0201
	h = 6			h = 9			h = 11		
	RMSPE Ratio	**t**	**P-Value**	**RMSPE Ratio**	**t**	**P-Value**	**RMSPE Ratio**	**t**	**P-Value**
SPF2	1.0114	-0.4790	0.3163	**0.9963**	0.1720	0.4318	**0.9971**	0.1463	0.4419
SPF11	1.0243	-0.3782	0.3529	**0.9821**	0.3398	0.3672	**0.9869**	0.2654	0.3955
SPF23	1.0745	-0.8135	0.2086	1.0028	-0.0401	0.4840	1.0031	-0.0487	0.4806
	h = 12			h = 18			h = 24		
	RMSPE Ratio	**t**	**P-Value**	**RMSPE Ratio**	**t**	**P-Value**	**RMSPE Ratio**	**t**	**P-Value**
SPF2	**0.9992**	0.0378	0.4849	1.0060	-0.3313	0.3704	1.0038	-0.3231	0.3735
SPF11	**0.9909**	0.1949	0.4228	1.0006	-0.0157	0.4937	**0.9922**	0.2518	0.4007
SPF23	1.0332	-0.0542	0.4784	1.0122	-0.2389	0.4057	**0.9977**	0.0484	0.4807

Notes: DMW test is constructed with HAC standard errors. RMSPEs lower than 1 favor survey-based forecasts. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead.

These results confirm two things. The first one, the SPF is no longer predicting accurate returns in the last subsample, there is no ratio below 1 at any horizon which is quite astonishing when we compare with the 2012M04-2018M05 subsample. And the second one, the appropriate benchmark for a survey-based forecast must be constructed by taking the closing price from the day before its release date, because it has a similar information set that SPF have when their results are published.

Table 5: Forecast accuracy of survey-based forecasts relative to the DRW+ at several forecasting horizons. 2012M04 – 2018M05 window.

	h = 1			h = 2			h = 3		
	RMSPE Ratio	T-stat	P-Value	RMSPE Ratio	T-stat	P-Value	RMSPE Ratio	T-stat	P-Value
SPF2	1.0348	-0.5681	0.2858	**0.9643**	0.8442	0.2007	**0.9579***	1.3111	0.0970
SPF11	1.1497	-1.5900	0.0581	**0.9818**	0.2569	0.399	**0.9233**	1.2213	0.1130
SPF23	1.3214	-2.1727	0.0165	1.0729	-0.6805	0.2492	**0.9862**	0.1477	0.4415
	h = 6			**h = 9**			**h = 11**		
	RMSPE Ratio	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**
SPF2	**0.9699***	1.2773	0.1029	**0.9837**	0.9314	0.1776	**0.9847**	0.8682	0.1943
SPF11	**0.9146***	1.4374	0.0776	**0.9121****	1.8872	0.0318	**0.9185****	1.7369	0.0436
SPF23	**0.9271**	0.9739	0.1668	**0.8969****	1.7513	0.0423	**0.8992****	1.7048	0.0466
	h = 12			**h = 18**			**h = 24**		
	RMSPE Ratio	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**
SPF2	**0.9821**	1.0341	0.1525	**0.9866**	0.836	0.2034	**0.9842***	1.4665	0.0744
SPF11	**0.9205****	1.7098	0.0462	**0.9259****	1.9342	0.0291	**0.9281*****	2.8127	0.0035
SPF23	**0.9002****	1.6806	0.0489	**0.9006****	1.9669	0.0271	**0.8957*****	2.7612	0.0040

Notes: DMW test is constructed with HAC standard errors. RMSPEs lower than 1 favor survey-based forecasts. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead.

Due to these results, from now on all future analysis are constructed based on the CLP closing price from the day before the survey is released, CLP+.

Table 6: Forecast accuracy of survey-based forecasts relative to the DRW+ at several forecasting horizons. 2018M06 – 2024M04 window.

	h = 1			h = 2			h = 3		
	RMSPE Ratio	T-stat	P-Value	RMSPE Ratio	T-stat	P-Value	RMSPE Ratio	T-stat	P-Value
SPF2	1.1038	-1.2222	0.1128	1.1233	-2.2394	0.0141	1.0912	-2.5497	0.0065
SPF11	1.4822	-2.9222	0.0023	1.3835	-3.2332	0.0009	1.2488	-2.4428	0.0085
SPF23	1.9140	-3.3549	0.0006	1.6724	-3.3264	0.0007	1.4145	-2.5477	0.0065
	h = 6			**h = 9**			**h = 11**		
	RMSPE Ratio	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**
SPF2	1.0296	-0.9226	0.1797	1.0031	-0.1099	0.4564	1.0050	-0.1827	0.4278
SPF11	1.0701	-0.8643	0.1952	1.0177	-0.2561	0.3993	1.0281	-0.4338	0.3329
SPF23	1.1348	-1.1697	0.1230	1.0552	-0.5930	0.2775	1.064	-0.7531	0.2269
	h = 12			**h = 18**			**h = 24**		
	RMSPE Ratio	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**	**RMSPE Ratio**	**T-stat**	**P-Value**
SPF2	1.0108	-0.4022	0.3444	1.0270	-0.9569	0.1709	1.0309	-1.6641	0.0503
SPF11	1.0361	-0.597	0.2762	1.0771	-1.5932	0.0578	1.0759	-2.4540	0.0083
SPF23	1.0678	-0.8665	0.1946	1.1226	-2.1556	0.0173	1.1255	-2.7590	0.0037

Notes: DMW test is constructed with HAC standard errors. RMSPEs lower than 1 favor survey-based forecasts. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead.

3.2 Mean Direction Accuracy Forecast Evaluation

Mean Direction is other common form to measure the accuracy of a series. Several papers show its importance like Nyberg and Pönka (2016) where they study directional predictability of monthly excess stock market returns for eleven markets. Pesaran and Timmermann (2002) shows how useful signs of stock returns are to take market timing decisions. Also, the ability to predict direction of change is quite important, because both the DRW and DRW+ has no sign at all, due to its nature of a no-change forecast.

To test this, we analyze Mean Direction Accuracy by taking the average of hit rate (HR), which is as follows:

$$HR_{t,h} = \begin{cases} 1 & if \quad (s_{t+h} - s_{t^+}) \cdot \left(s_{t^+}^{SPF}(h) - s_{t^+}\right) > 0 \\ 0 & if \quad (s_{t+h} - s_{t^+}) \cdot \left(s_{t^+}^{SPF}(h) - s_{t^+}\right) \leq 0 \end{cases} \quad (7)$$

We build hit rate based on CLP+, which is the closing price of the day before survey release for same reason as we state at the end of subsection 3.1. Then our hypotheses are:

$$H_0: E\left[HR_{t,h}\right] \leq 0.5$$
$$H_A: E\left[HR_{t,h}\right] > 0.5$$

In this case, we test whether the SPF outperforms a pure luck benchmark. We use the Gaussian t-statistic as applied in Cheung et al (2019) and evaluate it using a 1-tail test. This approach is justified because a low hit rate is equivalent to a high hit rate; for instance, a 20% hit rate is equivalent to an 80% hit rate if you simply take the opposite of what the forecast suggests. However, a low hit rate is not a desirable feature from a forecast, because it would reflect that the forecasts are completely lost.

Tables 7, 8, and 9 present the results for Mean Directional Accuracy in the full sample, the first subsample, and the second subsample, respectively.

Table 7: Directional Forecasting at several horizons. 2012M04 – 2024M04

Survey / Horizon	1	2	3	6	9	11	12	18	24
SPF2	51.7	50.0	48.3	45.7	46.0	48.9	47.0	50.8	46.7
SPF11	55.9*	55.6*	56.6*	51.4	56.2	59.3	56.0	56.3	53.3
SPF23	55.2	54.9	54.5	49.3	52.6	54.1	53.7	53.1	50.0

Notes: Gaussian t-statistic test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead.

Table 8: Directional Forecasting at several horizons. 2012M04 – 2018M05

Survey / Horizon	1	2	3	6	9	11	12	18	24
SPF2	56.8*	58.9**	55.6	56.5	54.5	59.4	57.1	52.6	56.9
SPF11	63.5***	63.0**	65.3***	56.5	62.1*	67.2**	63.5*	68.4**	76.5***
SPF23	60.8**	60.3**	62.5**	53.6	59.1	60.9	60.3	64.9*	72.5***

Notes: Gaussian t-statistic test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead.

Table 9: Directional Forecasting at several horizons. 2018M06 – 2024M04

Survey / Horizon	1	2	3	6	9	11	12	18	24
SPF2	46.5	40.8	40.8	35.2	38.0	39.4	38.0	49.3	39.4
SPF11	47.9	47.9	47.9	46.5	50.7	52.1	49.3	46.5	36.6
SPF23	49.3	49.3	46.5	45.1	46.5	47.9	47.9	43.7	33.8

Notes: Gaussian t-statistic test is constructed with HAC standard errors. Last available observation is from 2016M06 for 24 months ahead. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead.

Results in Table 7 shows that SPF has terrible accuracy in predicting direction of change. Presenting HR closed to 50%. With SPF11 at 11 months ahead being the highest of them all, with a 59.3% of HR.

On one hand, Table 8 is quite the opposite. SPF11 is the best in terms of predicting direction of change at almost all horizons. With 6 months ahead being the exception. Ranging from 63% for 2 months ahead to an astonishing 76.5% for 24 months ahead horizon.

On the other hand, Table 9 HR are awful. SPF2 displays quite low HR for 6 and 9 months ahead horizons, with 35.2% (64.8%) and 38% (62%). SPF23 reach the lowest with 33.8% (66.2%) for 24 months ahead. Even though we mention that a low HR is equivalent to a high one, this does not change the fact that survey-based forecast is completely lost in terms of direction of sign. Not being able to tell whether the returns will be positive or negative.

In this case survey has predictability over direction of change of CLP+ return, nevertheless the SPF hit rate has dropped. Besides that, we have an awful hit rate for full sample and the explanation is quite simple. If you want a good Mean Direction Accuracy forecast then you want a high hit rate or a lower one, but never a middle one. Then in the first subsample the hit rate is high while the second is low. Then if you average both subsamples the hit rate will converge into a 50 percent approximately.

3.3 Correlation Forecast Evaluation

When determining whether a series is predictable, most papers assert that a series is considered predictable if the forecast outperforms the DRW in terms of MSPE. Occasionally, predictability is evaluated using Mean Direction Accuracy, but almost never through the correlation between the target variable and its predictor, which seems contradictory. Diebold and Kilian (2001) define predictability as *"The extent of a series predictability depends on how much information the past conveys regarding future values of this series"* (Diebold and Kilian, 2001, p. 657). In other words, its correlation value. Pincheira and Hardy (2024) evaluate this feature both theoretically and empirically. They conclude that predictability and accuracy are different concepts when the forecast is not efficient according to the Mincer and Zarnowitz (1969) framework (henceforth MZ).

Therefore, we estimate correlation between the target variable, CLP+, and SPF+ at several horizons h =1, 2, 3, 6, 9, 11, 12, 18 and 24 months for full sample, first subsample and second subsample. The new target variable and forecast are as follows:

$$r_{t^+,t+h} = s_{t+h} - s_{t^+} \quad (8)$$
$$r_{t^+}^{SPF+} = s_{t^+}^{SPF} - s_{t^+} \quad (9)$$

We test correlation significance with a simple regression test just like Hansen (2022) with HAC standard errors:

$$r_{t^+,t+h} = \alpha + \beta \cdot r_{t^+}^{SPF+} + \mu_{t+} \ (10)$$

Then our null and alternative hypothesis are:

$$H_0: \beta = 0$$
$$H_A: \beta > 0$$

Our null hypothesis states that both variables have no correlation, in other words, they are linearly independent. Results are shown in Table 10, 11 and 12. We test β, however the entries present actual correlations values.

Table 10: Correlation between CLP+ and SPF at several horizons. 2012M04 – 2024M04 window.

	h = 1			h = 2			h = 3		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
SPF2	0.1764**	1.7395	0.042	0.1037	1.0561	0.1464	0.0739	1.0212	0.3089
SPF11	0.2003**	2.0695	0.0201	0.1542**	1.8737	0.0315	0.1497**	1.7534	0.0409
SPF23	0.1763**	1.8472	0.0334	0.1608**	1.8862	0.0307	0.1768**	1.7498	0.0412
	h = 6			h = 9			h = 11		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
SPF2	0.0859	0.7726	0.2212	0.1275	0.9226	0.1797	0.1134	0.8036	0.2121
SPF11	0.2106*	1.335	0.0930	0.2889*	1.5482	0.0630	0.2692*	1.304	0.0844
SPF23	0.2526*	1.5545	0.0623	0.3473**	1.9942	0.0250	0.3449**	1.9737	0.0262
	h = 12			h = 18			h = 24		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
SPF2	0.0997	0.6846	0.2479	0.0675	0.3737	0.3549	0.0205	0.1400	0.4445
SPF11	0.2565*	1.3279	0.0942	0.2400	1.1874	0.1195	0.2304*	1.3174	0.0960
SPF23	0.3473**	2.0259	0.0233	0.3500**	2.0035	0.0245	0.3624**	2.2732	0.0130

Notes: Hansen test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Correlations are actual correlations values.

Results are striking for the three samples. Correlations are, overall high, and statistically significant at several horizons, for short, medium, and long terms. In Table 10 we reject null for 18 out of 27 cases at least by 10% of significance. In Table 11 we reject 19 out of

27 cases at least by 10% of significance. And in Table 12 we reject null for 19 out of 27 entries. Also, correlations have increase in the second subsample at some horizons.

This shows that, even in turbulent times, the predictability of the SPF is still intact, as the correlation between the series remains high and statistically significant. In fact, the correlation even increases at certain horizons. However, despite retaining its predictability, the SPF accuracy has dropped.

Nevertheless, there may be a concern about persistence in long term horizons. Because as horizon increase so does its persistence. Then we could be facing a nonsense correlation. To see that is not the case, we plot the SPF forecast returns and CLP+ against time which is show in Figure 2. Series are scale up by 100.

Table 11: Correlation between CLP+ and SPF at several horizons. 2012M04 – 2018M05 window.

	h = 1			h = 2			h = 3		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
SPF2	0.1654*	1.2930	0.1001	0.2017*	1.6484	0.0518	0.2080*	1.5589	0.0618
SPF11	0.2308**	1.7844	0.0393	0.2730**	1.9239	0.0292	0.3333***	2.1874	0.0160
SPF23	0.225*	1.5077	0.068	0.2663*	1.6064	0.0563	0.2780*	1.5167	0.0669
	h = 6			**h = 9**			**h = 11**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
SPF2	0.1171	1.1895	0.1191	-0.0208	-0.1652	0.4346	-0.0486	-0.3652	0.3580
SPF11	0.2629**	1.9757	0.0260	0.2531**	1.6859	0.0481	0.2107*	1.2951	0.0997
SPF23	0.2538*	1.5635	0.0612	0.3051**	2.0026	0.0245	0.2925**	1.8462	0.0345
	h = 12			**h = 18**			**h = 24**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
SPF2	-0.0393	-0.3035	0.3812	-0.0811	-0.4511	0.3266	0.0508	0.2939	0.3848
SPF11	0.1865	1.0924	0.1392	0.2521	1.2215	0.1130	0.4459***	2.4802	0.0078
SPF23	0.2823**	1.7106	0.0458	0.3773**	2.3561	0.0106	0.6143***	4.8709	0.0000

Notes: Hansen test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Correlations are actual correlations values.

The behavior of SPF23 is very similar to the target variable in 2 months, however when we increase the horizons the magnitude of the returns is quite low compared to the CLP+. Nevertheless, we see that both series behave in a similar way. So, we can discard the fact that the autocorrelations of the series are causing that correlations are this high. In consequence, correlations make sense.

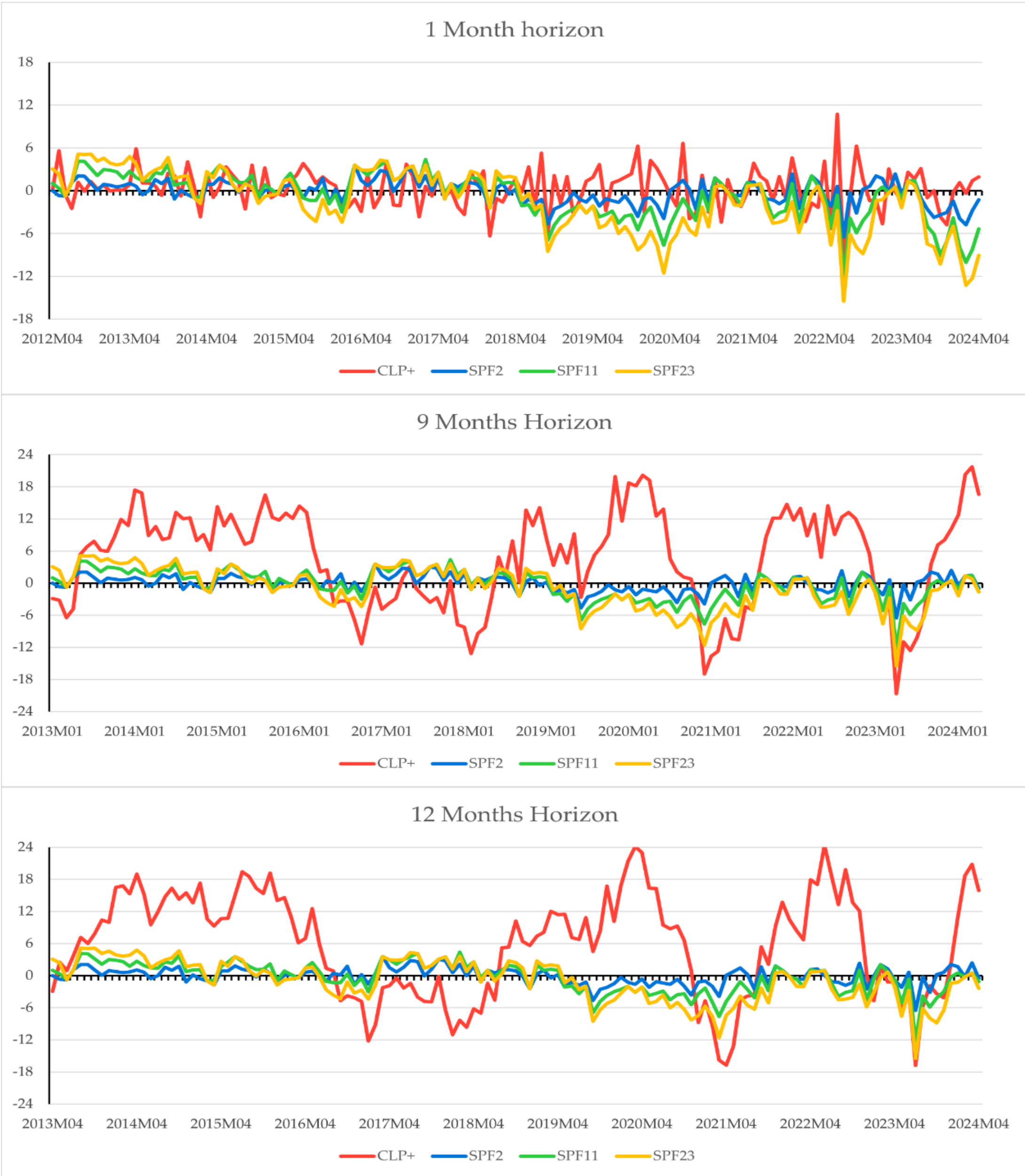
Table 12: Correlation between CLP+ and SPF at several horizons. 2018M06 – 2024M04 window.

	h = 1			h = 2			h = 3		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
SPF2	0.1899	1.1218	0.1329	0.0759	0.4611	0.3231	0.0559	0.4751	0.3181
SPF11	0.2500*	1.5165	0.0669	0.1875**	1.9061	0.0303	0.2064***	3.1471	0.0012
SPF23	0.2132*	1.3922	0.0841	0.2043*	1.6318	0.0536	0.2889***	4.0874	0.0001
	h = 6			**h = 9**			**h = 11**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
SPF2	0.1433	0.9369	0.1760	0.2699**	2.0798	0.0206	0.2523**	1.8518	0.0341
SPF11	0.3817***	2.5107	0.0072	0.5179***	4.1195	0.0001	0.4744***	3.2767	0.0008
SPF23	0.4727***	3.3148	0.0007	0.6004***	4.4505	0.0000	0.5750***	3.8640	0.0001
	h = 12			**h = 18**			**h = 24**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
SPF2	0.2140*	1.4950	0.0697	0.1319	0.7679	0.2226	-0.0404	-0.2540	0.4001
SPF11	0.4438***	3.1471	0.0012	0.3670*	1.3417	0.0920	0.1003	0.4326	0.3333
SPF23	0.5669***	4.0874	0.0001	0.4060**	2.1670	0.0168	0.2126	0.8740	0.1935

Notes: Hansen test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead. Correlations are actual correlations values.

To sum up, we find that the relative forecast accuracy of the survey has worsened for the second subsample. Yet its correlation are high and statistically significant. This is a situation labeled as the MSPE Paradox by Pincheira and Hardy (2024). The paradox occurs when the series with the highest correlation also has the highest MSPE. While the series with the lowest correlation presents the lowest MSPE as well. In their work they find that this happens when MZ efficiency conditions are not met. Then our forecast may be not efficient and/or present bias.

Figure 2: CLP+ and SPF forecasts for 1, 9 and 12 months.



Notes: Series are scale up by 100.

3.4. Range Coverage Analysis

Finally, we also evaluate the coverage range of the survey. The CBCH provides the values for the 10th and 90th percentiles of the respondents, which may be use as a sort of range. Nevertheless, these percentiles do not represent a range under any means, because the forecasters do not answer this question. Therefore, we are not evaluating under the Prediction Interval framework such as Chatfield (1993) or Christoffersen (1998), because the survey does not ask for an interval.

Then the goal of this analysis is to assess how reliable the survey is in terms of whether the actual value of the CLP falls between the 10th and 90th percentiles.

To do so, we once again calculate a Hit Rate based on the following formula:

$$Hit\ Rate\ Range = \begin{cases} 1 \ \ if \ \ r_{t^+,t+h} \geq r_{t^+,P10}^{SPF+} \ \ and \ r_{t^+,t+h} \leq r_{t^+,P90}^{SPF+} \\ 0 \ \ if \ \ r_{t^+,t+h} < r_{t^+,P10}^{SPF+} \ \ or \ r_{t^+,t+h} > r_{t^+,P90}^{SPF+} \end{cases}$$

These were calculated for horizons 1, 2, 6, 11, 12 and 24 months ahead. Results are display in Table 13, 14 and 15 for the full sample, first and second subsample, respectively. Entries shown percentage values.

Table 13: Range Coverage of SPF+ relative to the DRW+. 2012M04 – 2024M04

Survey / Horizon	1	2	6	11	12	24
SPF+2	56.6	41.0	14.3	10.4	11.2	5.7
SPF+11	71.0	61.8	36.4	20.0	22.4	11.5
SPF+23	52.1	71.5	56.4	39.3	36.6	17.2

Notes: Entries shown percentage values.

Table 14: Range Coverage of SPF+ relative to the DRW+. 2012M04 – 2018M05

Survey / Horizon	1	2	6	11	12	24
SPF+2	60.8	41.1	15.9	7.8	9.5	2.0
SPF+11	86.5	76.7	47.8	15.6	19.0	2.0
SPF+23	52.1	93.2	72.5	43.8	41.3	13.7

Notes: Entries shown percentage values.

Table 15: Range Coverage of SPF+ relative to the DRW+. 2018M06 – 2024M04

Survey / Horizon	1	2	6	11	12	24
SPF+2	52.1	40.8	12.7	12.7	12.7	8.5
SPF+11	54.9	46.5	25.4	23.9	25.4	18.3
SPF+23	52.1	49.3	40.8	35.2	32.4	19.7

Notes: Entries shown percentage values.

The results in Table 13 show that SPF+11 exhibits good range coverage for 1- and 2-months forecasts, while SPF+23 performs similarly for 2-months forecasts. However, for the other forecasting horizons, the coverage is poor. Additionally, both SPF+11 and SPF+23 show inadequate range coverage for the 11- and 24-month horizons, which is contradictory considering that these surveys are intended to target those specific horizons.

Table 14 results display something similar, however hit rate are higher for the three SPF at 1, 2 and 6 months ahead. Once again, this range coverage is awful when we reach the long-term horizons.

The results in Table 15 differ somewhat from the first subsample, as there is no good range coverage for the first three horizons, which was not an issue previously. These further highlights that SPF accuracy is diminished in the second subsample.

Anyways, the survey range coverage is bad overall. We see high range coverage for SPF+11 and SPF+23 at 1 and 2 forecasting horizons. This should surprise nobody because you expect that at longer horizons the returns should be higher. However, it surprises how poorly the range performs at the intended forecasting horizons for all three surveys.

4. Efficiency Analysis

Our previous results suggests that we may be facing a MSPE Paradox scenario. This happens when the forecast is not efficient in the MZ way. Which are two different conditions.

First is efficiency condition. Which is when forecast errors are orthogonal to the forecast. In other words, the information contained in the error term cannot be explained by the forecast. Second is the bias condition, which occurs when *"the forecast systematically understimated or overstimates levels of realizations…" (Mincer and Zarnowitz 1969).* Both are undesirable features, the first one because you are not using full information set. The second one increases MSPE.

Then we evaluate if SPF meet these conditions or not. For efficiency we run a simple regression test between forecast and forecast error. While for bias, we regress forecast error against a constant. We estimate both regressions with HAC standard errors.

4.1 Forecast Efficiency Evaluation

In our setup we do the following regression with HAC standard errors:

$$r^{SPF}_{t^+,t+h}(h) = \rho + \theta \cdot e^{SPF}{}_{t^+}(h) + u_{t+} \quad (11)$$

Where $e_{t^+}(h) = r_{t^+,t+h} - r^{SPF}_{t^+,t+h}(h)$.

Our hypotheses are as follows:

$$H_0: \theta = 0$$
$$H_A: \theta \neq 0$$

We test θ, but tables 13, 14 and 15 displays actual correlations.

The null hypothesis states that forecast is linearly independent with their forecast error. This mean that SPF prediction is not able to explain what is contain inside its error term, thus forecast is efficient. Results are display in Table 16, 17 and 18.

In Table 16, we see that efficiency condition is rejected in 9 out of 27 entries. This happens for all SPF in the first three horizons at 1% significance level. However, we cannot reject null hypothesis for rest of the entries. Also, these 9 cases are all negative correlations. In consequence the inefficiency condition increases the forecast accuracy.

While in Table 17. We reject for 8 out of 27 entries at least by 10% significance level. With the forecast being inefficient at 1 month ahead for the three surveys. Also, SPF23 proves to be an inefficient forecast for 1, 2, 3 and 24 months ahead. The short term and midterm entries where null is rejected are all negative correlations. In contrast, SPF23 at 24 months ahead is greater than 0.

In addition. In Table 18 we reject 12 out of 27 cases with almost same results as Table 14. Once again, the efficiency condition is not met for the first three horizons and is below 0. What is different from the first subsample is the fact that mid-term horizons are rejected for SPF11 and SPF23 at 9 months ahead. And SPF23 at 11 months ahead. All greater than 0.

Table 16: Forecast Inefficiency at several horizons. 2012M04 – 2024M04 window.

	h = 1			h = 2			h = 3		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
SPF2	-0.3971***	-4.0298	0.0001	-0.3343***	-4.072	0.0001	-0.2740***	-3.5840	0.0005
SPF11	-0.6829***	-7.9567	0.0000	-0.5707***	-6.5699	0.0000	-0.4525***	-4.5958	0.0000
SPF23	-0.8002***	-14.316	0.0000	-0.6962***	-10.369	0.0000	-0.5719***	-6.9886	0.0000
	h = 6			**h = 9**			**h = 11**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
SPF2	-0.1429	-1.1014	0.2727	-0.0425	-0.3122	0.7553	-0.0695	-0.4896	0.6252
SPF11	-0.1603	-0.9531	0.3422	-0.019	-0.0994	0.9209	-0.0431	-0.2221	0.8246
SPF23	-0.2363	-1.5676	0.1193	-0.0679	-0.4002	0.6896	-0.0674	-0.4004	0.6895
	h = 12			**h = 18**			**h = 24**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
SPF2	-0.0972	-0.6852	0.4944	-0.1050	-0.6189	0.5371	-0.0707	-0.5356	0.5932
SPF11	-0.0474	-0.2408	0.8101	-0.0463	-0.24	0.8107	0.0308	0.1768	0.8600
SPF23	-0.0618	-0.3623	0.7177	-0.0263	-0.1488	0.8820	0.0731	0.4109	0.6819

Notes: Test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%.

Overall, we find that there is not a major difference between the samples. SPF23 is not efficient for the first three months horizons in the three samples. Also, these correlations were all negative. Furthermore, the SPF forecast is more inefficient than in the first subsample, both at short and mid-term horizons, however at long term the null was not rejected.

Finally, our results may suggest that auto-efficiency condition is not a major factor behind the SPF inaccuracy. Of course, it affects the forecast performance, but when we compare the results between subsamples, we do not find a major difference.

Table 17: Forecast Inefficiency at several horizons. 2012M04 – 2018M05 window.

	h = 1			h = 2			h = 3		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
SPF2	-0.3474***	-3.4034	0.0011	-0.1525	-1.2711	0.2078	-0.0783	-0.5655	0.5735
SPF11	-0.4588***	-3.815	0.0003	-0.2192	-1.5645	0.1222	-0.0652	-0.4037	0.6877
SPF23	-0.6472***	-5.4208	0.0000	-0.4239***	-2.737	0.0078	-0.2987*	-1.7101	0.091
	h = 6			**h = 9**			**h = 11**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
SPF2	-0.2246**	-2.0872	0.0407	-0.1502	-1.127	0.2640	-0.1709	-1.275	0.2072
SPF11	-0.0532	-0.3491	0.7281	0.0575	0.3573	0.7221	-0.0009	-0.0054	0.9957
SPF23	-0.1497	-0.9667	0.3372	0.0212	0.1308	0.8963	0.0021	0.012	0.9905
	h = 12			**h = 18**			**h = 24**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
SPF2	-0.2030	-1.4741	0.1457	-0.1803	-1.0508	0.2980	0.0362	0.2094	0.8350
SPF11	0.0055	0.0299	0.9763	0.1253	0.5258	0.6012	0.4348**	2.0456	0.0463
SPF23	0.0266	0.1459	0.8845	0.203	0.8375	0.4060	0.5629**	2.5854	0.0128

Notes: Test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%.

4.2. Forecast Bias Evaluation

We evaluate bias condition, which is define as systematically under/overestimating the series value. This can be measured as the expected value of the forecast error. Where an expected value greater than 0 means that we are underestimating the actual return. While bias below 0 translates into an overestimation of the actual return.

We run a simple regression of forecast error against a constant with HAC covariance matrix estimation.

$$e^{SPF}{}_{t^{+}}(h) = \delta + \varepsilon_{t+h} \quad (12)$$

With the following hypotheses:

$$H_0: E[e^{SPF}{}_{t^{+}}(h)] = 0$$
$$H_A: E[e^{SPF}{}_{t^{+}}(h)] \neq 0$$

A null rejection implies that the forecast is bias. This bias translates into a more inaccurate forecast, which is an undesirable feature. In our case, bias can be interpreted as the percentage by which the forecast fell short or exceeded the actual return. As an example, if bias is 0.0277, then the forecast is underestimating the actual return by 2.77%.

Results are shown in Table 19, 20 and 21. Error terms are scale up by 100.

Table 18: Forecast Inefficiency at several horizons. 2018M06 – 2024M04 window.

	h = 1			h = 2			h = 3		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
SPF2	-0.3426**	-2.2454	0.0139	-0.3425***	-2.7378	0.0039	-0.2754***	-3.2113	0.0010
SPF11	-0.5402***	-4.2970	0.0000	-0.4583***	-4.4035	0.0000	-0.3187***	-3.1021	0.0014
SPF23	-0.6668***	-6.7634	0.0000	-0.5590***	-6.0416	0.0000	-0.3662***	-3.3497	0.0006
	h = 6			h = 9			h = 11		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
SPF2	-0.0135	-0.0820	0.4675	0.1039	0.7395	0.231	0.0449	0.3051	0.3806
SPF11	0.1216	0.6644	0.2543	0.2720*	1.5436	0.0636	0.1958	1.141	0.1288
SPF23	0.1417	0.7637	0.2238	0.2927*	1.6055	0.0564	0.2454*	1.4150	0.0807
	h = 12			h = 18			h = 24		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
SPF2	0.0041	0.0308	0.4878	-0.0772	-0.4364	0.3319	-0.1980	-1.1312	0.1309
SPF11	0.1555	0.9216	0.1799	-0.0628	-0.3221	0.3742	-0.2215	-0.8893	0.1884
SPF23	0.2096	1.1817	0.1206	-0.0122	-0.0605	0.4760	-0.2327	-0.8974	0.1863

Notes: Correlation test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead.

In full sample survey-based forecast is bias, this bias is positive and statistically significant at least by 10% for all horizons. Therefore, median forecaster is systematically underestimating the CLP exchange rate.

Full sample results shows that SPF is positive bias at all horizons for all three surveys, at least by 10%. In consequence SPF is systematically underestimating the CLP+ returns at all horizons. Also, the bias increases with its forecast horizons. This makes sense because our prediction values remain the same at all horizons. So, if we fell short at short-term, with major reason we fell short at mid and long term.

Table 19: Forecast Error Expected Value at several horizons. 2012M04 – 2024M04 window.

	h = 1			h = 2			h = 3		
	E[e]	T-stat	P-Value	E[e]	T-stat	P-Value	E[e]	T-stat	P-Value
SPF2	0.5493*	1.7254	0.0866	1.0052*	1.9635	0.0515	1.4311**	1.9762	0.0501
SPF11	1.076*	1.6764	0.0958	1.5069**	2.0058	0.0468	1.8981**	2.1134	0.0363
SPF23	1.8014**	2.0197	0.0453	2.2116**	2.2835	0.0239	2.5797**	2.3961	0.0179
	h = 6			h = 9			h = 11		
	E[e]	T-stat	P-Value	E[e]	T-stat	P-Value	E[e]	T-stat	P-Value
SPF2	3.2250**	2.2294	0.0274	4.5969**	2.4607	0.0151	5.4573***	2.6307	0.0095
SPF11	3.5928**	2.4202	0.0168	4.8958***	2.6438	0.0092	5.7559***	2.8206	0.0055
SPF23	4.254***	2.7582	0.0066	5.5315***	2.9992	0.0032	6.3964***	3.193	0.0018
	h = 12			h = 18			h = 24		
	E[e]	T-stat	P-Value	E[e]	T-stat	P-Value	E[e]	T-stat	P-Value
SPF2	5.8948***	2.7219	0.0074	8.6376***	3.2980	0.0013	11.5328***	3.9312	0.0001
SPF11	6.1957***	2.9153	0.0042	8.8773***	3.4720	0.0007	11.6131***	4.061	0.0001
SPF23	6.841***	3.2958	0.0013	9.4926***	3.8603	0.0002	12.1202***	4.4346	0.0000

Notes: Expected value test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Forecast errors are scale up by 100.

However, first subsample results are drastically different. We reject just 9 out of 27 entries. Also, some of the biases are negative, like SPF11 and SPF23 at 1 month ahead. Which means that the forecast is overestimating the actual return. Then we do not reject until 12, 18 and 24 months ahead entries.

Which is a major contrast with the second subsample. Once again, all entries present strong and positive bias. Null is rejected at least by 5% of significance level. Besides that, a major finding is that the magnitude of the bias has increase for SPF in general at all horizons. So, it is not just that the survey-based forecast is systematically underestimating actual returns. But also, that the percentage by which it falls short is higher. This leads into a higher MSPE and in consequence, a major inaccuracy.

Table 20: Forecast Error Expected Value at several horizons. 2012M04 – 2018M05 window.

	h = 1			h = 2			h = 3		
	E[e]	T-stat	P-Value	E[e]	T-stat	P-Value	E[e]	T-stat	P-Value
SPF2	-0.1052	-0.3355	0.7382	0.2296	0.4286	0.6695	0.4527	0.5735	0.5681
SPF11	-0.9361**	-2.4074	0.0186	-0.6126	-1.0521	0.2963	-0.3897	-0.4787	0.6336
SPF23	-1.0056*	-1.7173	0.0902	-0.6721	-0.893	0.3748	-0.4387	-0.4583	0.6482
	h = 6			**h = 9**			**h = 11**		
	E[e]	**T-stat**	**P-Value**	**E[e]**	**T-stat**	**P-Value**	**E[e]**	**T-stat**	**P-Value**
SPF2	1.4193	1.0372	0.3034	2.7666	1.3163	0.1928	3.7950	1.5654	0.1226
SPF11	0.5877	0.5900	0.5572	1.9211	1.0069	0.3178	2.9354	1.2955	0.1999
SPF23	0.5719	0.5734	0.5683	1.9282	1.0347	0.3047	2.9430	1.3299	0.1884
	h = 12			**h = 18**			**h = 24**		
	E[e]	**T-stat**	**P-Value**	**E[e]**	**T-stat**	**P-Value**	**E[e]**	**T-stat**	**P-Value**
SPF2	4.4077*	1.6873	0.0967	8.0035**	2.1271	0.0379	11.7305**	2.4869	0.0163
SPF11	3.5226	1.4354	0.1563	7.1299*	1.9911	0.0515	10.8556**	2.4357	0.0185
SPF23	3.5184	1.4796	0.1441	7.0861**	2.073	0.0429	10.8209**	2.6072	0.0121

Notes: Expected value test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Forecast errors are scale up by 100.

Our findings reveal that bias explains a lot of the SPF inaccuracy in the second subsample[2]. First, we find that SPF is strongly positive bias at all horizons. Second, we reject null hypotheses for all entries at least by 5%. Third, bias magnitude has increase in the last subsample, leading into a more inaccurate forecast.

In summary, the SPF presents little bias in the first subsample. However, in the second subsample we find that the survey-based forecast underestimates CLP+ returns in a consistent way. Furthermore, someone could think that this is caused by exogenous events that has affected global and Chilean economy. Nevertheless, these events occurred years ago. So, the remaining questions are, what happened with the survey-

[2] Due to these results, we may think that forecasting with the 90th percentile of the survey could be a reasonable option to fix this bias problem. Therefore, we test both MSPE and direction of change. Results are in Appendix C.

based forecast in the last subsample? Is due to a structural change in the CLP exchange rate? Does the SPF have not been able to fully adapt its expectations? Can we optimize this forecast to obtain better results?

Table 21: Forecast Error Expected Value at several horizons. 2018M06 – 2024M04 window.

	h = 1			h = 2			h = 3		
	E[e]	T-stat	P-Value	E[e]	T-stat	P-Value	E[e]	T-stat	P-Value
SPF2	1.0878***	2.7433	0.0077	1.6001**	2.3465	0.0217	2.1292**	2.1936	0.0315
SPF11	3.0740***	4.6827	0.0000	3.5455***	4.0804	0.0001	4.0127***	3.6408	0.0005
SPF23	4.7391***	5.5729	0.0000	5.1766***	5.1393	0.0000	5.6049***	4.7205	0.0000
	h = 6			**h = 9**			**h = 11**		
	E[e]	**T-stat**	**P-Value**	**E[e]**	**T-stat**	**P-Value**	**E[e]**	**T-stat**	**P-Value**
SPF2	3.5711**	2.1401	0.0358	5.1586**	2.3607	0.0210	6.1306**	2.5017	0.0147
SPF11	5.3564***	3.0597	0.0031	6.7919***	3.1317	0.0025	7.6779***	3.2353	0.0018
SPF23	6.9103***	3.9630	0.0002	8.2493***	3.9620	0.0002	9.0711***	4.0597	0.0001
	h = 12			**h = 18**			**h = 24**		
	E[e]	**T-stat**	**P-Value**	**E[e]**	**T-stat**	**P-Value**	**E[e]**	**T-stat**	**P-Value**
SPF2	6.5543**	2.5852	0.0118	9.0482***	3.4759	0.0009	11.8295***	4.7881	0.0000
SPF11	8.1015***	3.3011	0.0015	10.4401***	4.0291	0.0001	12.8175***	5.0269	0.0000
SPF23	9.4723***	4.1156	0.0001	11.7077***	4.7943	0.0000	13.8537***	5.4920	0.0000

Notes: Expected value test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Forecast errors are scale up by 100. Last available observation is from 2016M06 for 24 months ahead.

5. Bias Adjusted Forecast

We find that survey-based forecast is positive bias at all horizons for the last subsample. This may suggest that bias is the major factor behind forecast inaccuracy. Yet it can predict CLP+, presenting high and statistically significant correlations.

This leads into how do we value this forecast? Because it can predict CLP+ returns, but in an imprecise way. Mincer and Zarnowitz states that a biased forecast may be more accurate if its prediction incorporates its bias. For example, a moving average rolling window of pasts forecast errors.

Which is possible because SPF forecast is bias. Our approach is simple, we predict CLP+ with SPF and calculate its average forecast error with a moving average rolling window from 2012M04 to 2018M05 for all horizons. This means that when we predict CLP+ returns for 2018M06 at 24 months ahead our last observation is from 2016M06. Then this is a real time prediction exercise.

These new Bias Adjusted (BA) SPF+ return and forecast error are as follows:

$$r_{t^+}^{SPF+MA}(h) = s_{t^+}^{SPF+} - s_{t^+} + \gamma_t \quad (13)$$
$$e_{t^+}^{SPF+MA} = s_{t+h} - s_{t^+}^{SPF+} - \gamma_t \quad (14)$$

Where $\gamma_t = \frac{\sum_{i=1}^{t}\left(e_{t^+,t+h}^{SPF+}\right)_i}{t}$ represents forecast error moving average.

We evaluate relative MSPE and Mean Direction Accuracy at several horizons for the second subsample, with HAC standard errors. Results are display in Table 22 and 23.

Table 22: Forecast accuracy of the BA forecast relative to the DRW+ at several forecasting horizons. 2018M06 – 2024M04 window.[3]

	h = 1			**h = 2**			**h = 3**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
BA SPF+2	1.0778	-0.9440	0.1742	1.0803	-1.2838	0.1017	1.0457	-1.0792	0.1421
BA SPF+11	1.3522	-2.7051	0.0043	1.2507	-3.0131	0.0018	1.1316	-2.0851	0.0203
BA SPF+23	1.6195	-2.9979	0.0019	1.4095	-2.7494	0.0038	1.1948	-1.7346	0.0436
	h = 6			**h = 9**			**h = 11**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
BA SPF+2	**0.9788**	0.4226	0.3369	**0.9217**	1.2744	0.1033	**0.8996***	1.4433	0.0767
BA SPF+11	**0.9701**	0.556	0.2900	**0.8987****	1.8628	0.0333	**0.8896****	1.8379	0.0351
BA SPF+23	**0.9662**	0.4470	0.3281	**0.8764****	1.858	0.0337	**0.8674****	1.9286	0.0289
	h = 12			**h = 18**			**h = 24**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
BA SPF+2	**0.9023***	1.4794	0.0717	**0.915**	0.7288	0.2343	1.0342	-0.1816	0.4282
BA SPF+11	**0.895****	1.8531	0.0340	**0.9481**	0.4412	0.3302	1.0688	-0.372	0.3555
BA SPF+23	**0.8691****	2.0521	0.0219	**0.9371**	0.5300	0.2989	1.0595	-0.3175	0.3759

[3] Results against the RWD+ are shown in Table B.1, Appendix B. As expected from a tougher benchmark, RMSPE ratios are higher than with the DRW+.

Notes: DMW test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Ratios below 1 are in bold. Last available observation is from 2016M06 for 24 months ahead.

Table 22 show striking results, we reject 8 out of 27 cases. Our new forecast outperforms the DRW+ at 9, 11 and 12 months ahead. Also, the RMSPE ratios are lower than before, in all the entries. The BA SPF is more accurate than SPF, showing that bias place an important role on why survey-based forecast accuracy dropped in the second subsample. This improvement increases as horizons does because bias is higher at longer horizons, however for long term like 18 and 24 months ahead we are not able to reject the null.

The we evaluate Mean Direction Accuracy at several horizons for our new improve forecast in the same way as before, by performing a Gaussian t-statistic constructed with HAC standard errors and compare to a pure luck benchmark. Results are shown in Table 23.

Table 23: Directional Forecasting at several horizons for BA SPF. 2018M06 – 2024M04

Survey / Horizon	**1**	**2**	**3**	**6**	**9**	**11**	**12**	**18**	**24**
BA SPF+2	45.1	39.4	47.9	67.6**	76.1***	73.2**	71.8***	73.2***	63.4
BA SPF+11	50.7	46.5	46.5	47.9	63.4*	60.6	59.2	71.8**	64.8
BA SPF+23	49.3	43.7	43.7	57.7	67.6**	66.2**	64.8**	70.4**	63.4

Notes: Gaussian t-statistic test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Values are display in percentage.

Like Table 19 results, Mean Direction Accuracy is not great for first three months horizons. However, as horizons increases, the hit rate does as well. We highlight SPF+MA2 for achieving significant hit rates at 6, 9, 11, 12, and 18 months, with an impressive 76.1% accuracy for the 9-month horizon.

Finally, we evaluate correlations between the BA SPF+ and the CLP+. We expect that correlations have weakened. Results are display in Table 24; entries are actual correlation values.

For Table 24 results we reject the null for 10 out of 27 entries at least by 10% at short-, mid- and long-term horizons. Nevertheless, these values are lower than those shown in Table 12 for all entries. Also, at some horizons the correlations turn into negative values, BA SPF2+ and BA SPF23+ at 6 and 24 months ahead respectively just to name a few. Values that were quite rare before.

Table 24: Correlation between CLP+ and the BA SPF+ at several horizons. 2018M06 – 2024M04 window.

	h = 1			h = 2			h = 3		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
BA SPF+2	0.1706	1.0337	0.1524	0.0414	0.2357	0.4072	-0.0072	-0.0552	0.4781
BA SPF+11	0.2178*	1.4067	0.0819	0.1409	0.9019	0.1851	0.1407	1.1505	0.1269
BA SPF+23	0.1784	1.1653	0.1239	0.1529	1.0712	0.1438	0.2211**	1.7596	0.0414
	h = 6			h = 9			h = 11		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
BA SPF+2	-0.0146	-0.0951	0.4622	0.1040	0.7400	0.2309	0.0929	0.5935	0.2774
BA SPF+11	0.2493*	1.5708	0.0603	0.3378**	2.2238	0.0147	0.2766*	1.5477	0.0631
BA SPF+23	0.3628**	2.2943	0.0124	0.4626***	3.1762	0.0011	0.4105***	2.4664	0.0080
	h = 12			h = 18			h = 24		
	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value	Correlation	T-stat	P-Value
BA SPF+2	0.0574	0.3430	0.3663	-0.0436	-0.1858	0.4266	-0.1676	-0.7991	0.2135
BA SPF+11	0.2345	1.2655	0.1049	0.0325	0.1321	0.4476	-0.1001	-0.4680	0.3206
BA SPF+23	0.3838**	2.2672	0.0132	0.1455	0.5884	0.2791	-0.0272	-0.1261	0.4500

Notes: Hansen test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead. Correlations are actual correlations values.

6. Concluding Remarks

In this paper we re-evaluate if the survey-based forecasts of the Chilean exchange rate (SPF) predict Chilean Peso (CLP) returns at several forecasting horizons for the 2018M06 – 2024M04 period.

We found that its correlation with the target variable is high and significant; however, when compared to the Driftless Random Walk (DRW) and our more competitive benchmark, DRW+, it is consistently outperformed in terms of Mean Squared Prediction Error (MSPE). This may suggest a scenario consistent with the Pincheira and Hardy (2024) MSPE paradox.

We found that the SPF has dropped its accuracy under Mean Squared Prediction Error relative to the Driftless Random Walk (DRW) and the DRW+ in the June 2018 to April 2024 period. Mean Direction Accuracy has decreased too. However, the predictability remains, as correlation between the SPF and the target variable is high and statistically significant at several horizons.

This is consistent with the Pincheira and Hardy (2024) paradox scenario. Where the variable with the highest correlation has the highest MSPE too. According to their work, this occurs when the forecast does not met Mincer and Zarnowitz (1969) efficiency conditions.

Therefore, we evaluate these conditions and found that the SPF presents a positive bias at all horizons in the 2018M06 – 2024M04 period. This implies a systematically underestimation of the CLP actual returns.

To address this issue, we propose a new optimized forecast (SPF+MA) that corrects for bias. Our analysis shows that SPF+MA outperforms the DRW+ at mid- and long-term horizons. Additionally, Mean Direction Accuracy is significantly improved, particularly for these same horizons.

Our findings are significant for three key reasons. First, they challenge the long-held belief that exchange rates are not predictable. Second, we propose a more appropriate benchmark for evaluating survey-based forecasts. Third, we address the bias issue with a simple real-time approach.

Lastly, future research could focus on analyzing structural changes in the CLP due to past exogenous events such as the Chilean Social Outbreak, COVID-19, the Russia-Ukraine war, and two national referendums. These events may explain the major shifts in SPF bias. Another potential research avenue is the analysis of individual forecasters' expectations using a panel regression, with data available upon request from the Central Bank of Chile. Lastly, Engel and West (2005) find that exchange rates under certain conditions presents a near-random walk behavior. In consequence, the survey may also present a near-random walk behavior, then it may be easier to predict the SPF rather than the CLP.

Appendix A:

As shown in Figure 1, the CLP exhibits a notable depreciation trend in two distinct periods: from January 2014 to April 2016, and from October 2019 to April 2024. Given this trend, we may argue that a Random Walk with Drift (RWD) serves as a more appropriate benchmark than DRW, as the captures this depreciation trend. For instance, Meese and Rose (1991) examine the empirical relationship between nominal exchange rates and macroeconomic fundamentals, finding some evidence favoring nonlinear models, although not to a significant level.

As stated in subsection 3.1, the correct method for testing a survey-based forecast is to compare it with the CLP closing price from the day before the survey's release. Therefore, we use CLP+ instead of CLP. The constant (σ_{t^+}) is calculated using a moving average of past returns, focusing on the June 2018 to April 2024 period. Then the forecast and forecast error of the Random Walk with Drift Plus (RWD+) are:

$$r_{t^+}^{RWD+}(h) = \sigma_{t^+} + s_{t^+} - s_t$$
$$e_{t^+}^{RWD+}(h) = s_{t+h} - \sigma_{t^+} - s_{t^+}$$

Results are display in Table A.1

Table A.1: Forecast Accuracy of the SPF+ relative to the RWD+. 2018M06 – 2024M04 window.

	h = 1			h = 2			h = 3		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
SPF+2	1.1017	-1.1322	0.1307	1.1350	-2.0373	0.0227	1.1112	-1.8963	0.0310
SPF+11	1.4793	-2.8604	0.0028	1.3979	-3.0405	0.0017	1.2717	-2.1304	0.0183
SPF+23	1.9103	-3.3210	0.0007	1.6898	-3.2194	0.0010	1.4404	-2.3502	0.0108
	h = 6			**h = 9**			**h = 11**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
SPF+2	1.0296	-0.8183	0.208	1.0717	-0.7170	0.2379	1.1075	-0.9376	0.1758
SPF+11	1.0701	-0.8105	0.2102	1.0873	-0.3368	0.3686	1.1330	-0.8529	0.1983
SPF+23	1.1348	-1.0495	0.1488	1.1274	-1.2703	0.1041	1.1725	-0.9848	0.1640
	h = 12			**h = 18**			**h = 24**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
SPF+2	1.2240	-2.0939	0.0199	1.1664	-1.1297	0.1312	1.0782	-0.4189	0.3383
SPF+11	1.1485	-0.9699	0.1677	1.2233	-1.4281	0.0788	1.1253	-0.6822	0.2487
SPF+23	1.1836	-1.0834	0.1412	1.2749	-1.7684	0.0406	1.1772	-0.9665	0.1685

Notes: DMW test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Ratios below 1 are in bold. Last available observation is from 2016M06 for 24 months ahead.

Table A.1 results show that we cannot reject any case at all. Besides that, no RMSPE ratio is below 1. In consequence, the SPF is not able to outperform the RWD+ at any horizon in the second subsample. Also, the ratios are higher when we compare it to the DRW+ results in table 6.

This finding differs from Rossi's (2013) survey, where typically, the random walk with drift forecasts is worse than without drift. However, this may hold true just for this scenario, where the trend is notorious. Because in general, the DRW forecast tend to be more accurate than with drift. For example, Chinn (1991) and Chinn and Meese (1995) find an opposite result than Meese and Rose (1991), where the DRW outperforms the non-linear models at short term horizons.

Appendix B

Table B.1: Forecast Accuracy of the BA SPF+ relative to the RWD+. 2018M06 – 2024M04 window.

	h = 1			h = 2			h = 3		
	RMSPE ratio	T-stat	P-Value	RMSPE ratio	T-stat	P-Value	RMSPE ratio	T-stat	P-Value
BA SPF+2	1.0758	-0.8744	0.1924	1.0915	-1.5306	0.0652	1.0649	-1.786	0.0392
BA SPF+11	1.3497	-2.6082	0.0055	1.2638	-2.6208	0.0054	1.1523	-1.8165	0.0368
BA SPF+23	1.6164	-2.9372	0.0022	1.4242	-2.7294	0.0040	1.2167	-1.5987	0.0572
	h = 6			**h = 9**			**h = 11**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
BA SPF+2	1.0132	-0.3849	0.3507	**0.9847**	0.4682	0.3206	**0.9914**	0.2369	0.4067
BA SPF+11	1.0043	-0.0517	0.4795	**0.9601**	0.4620	0.3227	**0.9803**	0.2115	0.4165
BA SPF+23	1.0002	-0.0016	0.4994	**0.9363**	0.5576	0.2894	**0.9559**	0.3695	0.3564
	h = 12			**h = 18**			**h = 24**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
BA SPF+2	1.0002	-0.0059	0.4977	1.0392	-0.8695	0.1938	1.0816	-2.8355	0.003
BA SPF+11	**0.992**	0.0878	0.4651	1.0767	-0.912	0.1824	1.1179	-2.4701	0.008
BA SPF+23	**0.9633**	0.3222	0.3741	1.0642	-0.7327	0.2331	1.1082	-1.8476	0.0344

Notes: DMW test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Ratios below 1 are in bold. Last available observation is from 2016M06 for 24 months ahead.

As expected from a tougher benchmark, the results are worse than those shown in Table 22. We reject the null for 0 out of 27 entries, which are similar results before the adjustment. Also, the RSMPE ratios has increase, and in some cases, they shift from below 1 to above 1. For example, with the 6 months ahead horizon.

Nevertheless, the results of the BA forecast are better than without it. Also, the fact that the RWD+ is tougher than the DRW+ differs from what Rossi's (2013) survey says. That the tougher benchmark is the Random Walk without Drift. This does not hold true for trend series, just like Figure 1 shows. Then a Random Walk with Drift may be a more appropriate benchmark for capturing the underlying trend.

Appendix C

As shown in Figure 1 and Table 21, the SPF median presents a strongly positive bias. This implies that is systematically underestimating the CLP+ actual returns. Then, the CBCH also publishes the 90th percentile of the forecasters (SPF+ P90). So, it would be reasonable to think that 90th percentile forecast is closer to the CLP actual value. Therefore, we estimate SPF+ P90 MSPE accuracy relative to the DRW+ for the June 2018 to April 2024 period. Results are displayed in Table C.1. We also evaluate Directional Mean Accuracy for the SPF+ P90. Results are displayed in Table C.2.

Table C.1: Forecast Accuracy of the SPF+ P90 relative to the DRW+. 2018M06 – 2024M04 window.

	h = 1			h = 2			h = 3		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
SPF+2 P90	1.2278	-1.9218	0.0293	1.1451	-1.2459	0.1085	1.0779	-0.8403	0.2018
SPF+11 P90	1.2524	-2.2369	0.0142	1.1270	-1.2358	0.1103	1.0160	-0.1932	0.4237
SPF+23 P90	1.3395	-2.4653	0.0081	1.1655	-1.4285	0.0788	1.0148	-0.1689	0.4332
	h = 6			**h = 9**			**h = 11**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
SPF+2 P90	1.0123	-0.1805	0.4287	**0.9665**	0.6292	0.2656	**0.9563**	0.8554	0.1976
SPF+11 P90	**0.9495**	0.7997	0.2133	**0.9002****	2.0108	0.0241	**0.8912****	2.1357	0.0181
SPF+23 P90	**0.9156***	1.3305	0.0938	**0.8553*****	2.9705	0.0020	**0.8530*****	2.9241	0.0023
	h = 12			**h = 18**			**h = 24**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
SPF+2 P90	**0.9482**	1.0991	0.1377	**0.9364****	1.7611	0.0413	**0.9114*****	3.2644	0.0008
SPF+11 P90	**0.8894****	2.257	0.0135	**0.8910****	2.0367	0.0227	**0.8623*****	2.7559	0.0037
SPF+23P90	**0.8551*****	2.9463	0.0022	**0.8485****	2.1824	0.0162	**0.8287*****	2.4569	0.0082

Notes: DMW test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Ratios below 1 are in bold. Last available observation is from 2016M06 for 24 months ahead.

On one hand, Table C.1 display incredible results. We reject the null for 13 out of 27 entries, which is a higher number than the BA SPF+ forecast results in Table 22. The SPF+P90 forecast outperforms the DRW+ at mid- and long-term horizons, highlighting the 23-month survey. With the 90th percentile we find similar results as those displayed in the first subsample.

On the other hand. Table C.2 show similar results compare to Table 23; however, the BA forecast presents greater Mean Direction Accuracy, in terms of Hit Rate percentage. Also, we reject just 8 out of 27 entries, while for the BA forecast is 11 out 27. Nevertheless, the results are amazing, because at mid- and long-term horizons reaches high and significant Hit Rates. Such as SPF+2 at 9, 11, 12 and 18 forecasting horizons, with significant hit rates above 70%. In consequence, the 90th percentile is a viable alternative when positive bias is present.

Table C.2: Directional Forecasting at several horizons for SPF+ P90. 2018M06 – 2024M04

Survey / Horizon	1	2	3	6	9	11	12	18	24
SPF+2 P90	42.3	43.7	47.9	63.4*	74.6***	73.2*	71.8*	71.8***	62.0
SPF+11 P90	49.3	47.9	46.5	49.3	64.8*	60.6	60.6	70.4**	63.4
SPF+23 P90	52.1	50.7	49.3	49.3	57.7	62.0	59.2	69.0**	63.4

Notes: Gaussian t-statistic test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Values are display in percentage.

Table C.3: Correlation between CLP+ and the SPF+ P90 at several horizons. 2018M06 – 2024M04 window.

	h = 1			**h = 2**			**h = 3**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
BA SPF+2	0.1023	0.8754	0.1921	-0.0344	-0.2203	0.4131	-0.1066	-0.8764	0.1919
BA SPF+11	0.0983	0.9785	0.1656	0.0565	0.5039	0.3080	0.1070	0.8750	0.1923
BA SPF+23	0.1025	0.9425	0.1746	0.0962	0.8636	0.1954	0.1793*	1.4278	0.0789
	h = 6			**h = 9**			**h = 11**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
BA SPF+2	-0.1880*	-1.3418	0.0920	-0.0748	-0.5315	0.2984	-0.0891	-0.6191	0.2689
BA SPF+11	0.1106	0.7453	0.2293	0.2508*	1.6306	0.0537	0.2384*	1.2943	0.0999
BA SPF+23	0.2990**	1.6666	0.0500	0.3998***	2.8070	0.0032	0.3649**	2.2690	0.0132
	h = 12			**h = 18**			**h = 24**		
	Correlation	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**	**Correlation**	**T-stat**	**P-Value**
BA SPF+2	-0.0444	-0.3175	0.3759	-0.1194	-0.7690	0.2222	-0.0696	-0.4380	0.3314
BA SPF+11	0.2513*	1.3548	0.0899	0.1188	0.5387	0.2959	0.1120	0.4992	0.3096
BA SPF+23	0.3576**	2.1940	0.0158	0.2515	1.1961	0.1178	0.1788	0.7346	0.2325

Notes: Hansen test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Last available observation is from 2016M06 for 24 months ahead. Correlations are actual correlations values.

In Table C.3 we find that SPF+ P90 has strong and significant correlation with the target variable at short- and mid- term horizons. We reject 9 out 27 entries.

However, these values are lower than those with the SPF median, not only that, but we even find significant negative correlations in SPF2 at 6 months ahead. These results furthers deep the MSPE Paradox case that we have found with the median. Nevertheless, the 90th percentiles does predict the CLP+ returns.

Table C.4: Forecast Accuracy of the SPF+ P90 relative to the RWD+. 2018M06 – 2024M04 window.

	h = 1			h = 2			h = 3		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
SPF+2 P90	1.2255	-1.9379	0.0283	1.1571	-1.4913	0.0702	1.0976	-1.3244	0.0948
SPF+11 P90	1.2500	-2.2238	0.0147	1.1388	-1.5096	0.0678	1.0346	-0.5027	0.3084
SPF+23 P90	1.3369	-2.5349	0.0067	1.1776	-1.6587	0.0508	1.0333	-0.4240	0.3364
	h = 6			**h = 9**			**h = 11**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
SPF+2 P90	1.0480	-1.0883	0.1401	1.0326	-0.6971	0.2440	1.0538	-0.9430	0.1744
SPF+11 P90	**0.9829**	0.3254	0.3729	**0.9617**	0.6552	0.2572	**0.9821**	0.2584	0.3984
SPF+23 P90	**0.9478**	0.8725	0.1929	**0.9138**	1.2968	0.0995	**0.9400**	0.8120	0.2098
	h = 12			**h = 18**			**h = 24**		
	RMSPE ratio	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**	**RMSPE ratio**	**T-stat**	**P-Value**
SPF+2 P90	1.1591	-2.7959	0.0033	1.0634	-0.5739	0.2839	**0.9533**	0.2712	0.3935
SPF+11 P90	**0.9859**	0.2000	0.4210	**1.0119**	-0.1273	0.4495	**0.9019**	0.6247	0.2671
SPF+23P90	**0.9479**	0.7013	0.2427	**0.9636**	0.4667	0.3211	**0.8668**	0.9074	0.1836

Notes: DMW test is constructed with HAC standard errors. * Significance at 10%, ** significance at 5%, ***significance at 1%. Ratios below 1 are in bold. Last available observation is from 2016M06 for 24 months ahead.

References:

1. Brown, P. P., & Hardy, N. (2024). The mean squared prediction error paradox. Journal of Forecasting, 43(6), 2298–2321. https://doi.org/10.1002/for.3129
2. Capistrán, C. (n.d.). LAS EXPECTATIVAS MACROECONÓMICAS DE LOS ESPECIALISTAS.
3. Cheung, Y.-W., Chinn, M. D., Pascual, A. G., & Zhang, Y. (2019). Exchange rate prediction redux: New models, new data, new currencies. Journal of International Money and Finance, 95, 332–362. https://doi.org/10.1016/j.jimonfin.2018.03.010
4. Engel, C., Mark, N. C., & West, K. D. (2015). Factor Model Forecasts of Exchange Rates. Econometric Reviews, 34(1–2), 32–55. https://doi.org/10.1080/07474938.2014.944467
5. Ince, O., & Molodtsova, T. (2017). Rationality and forecasting accuracy of exchange rate expectations: Evidence from survey-based forecasts. Journal of International Financial Markets, Institutions and Money, 47, 131–151. https://doi.org/10.1016/j.intfin.2016.11.002
6. Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies. Journal of International Economics, 14(1–2), 3–24. https://doi.org/10.1016/0022-1996(83)90017-X
7. Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55(3), 703–708. https://doi.org/10.2307/1913610
8. Nyberg, H., & Pönkä, H. (2016). International sign predictability of stock returns: The role of the United States. Economic Modelling, 58, 323–338. https://doi.org/10.1016/j.econmod.2016.06.013
9. Pincheira-Brown, P., & Neumann, F. (2020). Can we beat the Random Walk? The case of survey-based exchange rate forecasts in Chile. Finance Research Letters, 37, 101380. https://doi.org/10.1016/j.frl.2019.101380
10. Rossi, B. (2013). Exchange Rate Predictability. Journal of Economic Literature, 51(4), 1063–1119. https://doi.org/10.1257/jel.51.4.1063
11. Pincheira, P. & Hardy, N. (2024). More Predictable than ever, with the worst MSPE ever (Working paper upon request). University Adolfo Ibañez and Diego Portales University.
12. Pesaran, M. H., & Timmermann, A. (2002). Market timing and return prediction under model instability. Journal of Empirical Finance, 9(5), 495–510. https://doi.org/10.1016/S0927-5398(02)00007-5
13. Pedersen, M. (2010). Una nota introductoria a la encuesta de Expectativas Económicas. Estudios Económicos Estadísticos N°82, Banco Central de Chile.
14. Diebold, F. X., and L. Kilian (2001). Measuring Predictability: Theory and Macroeconomic Applications. Journal of Applied Econometrics, 16: 657–669 (2001).

15. Meese, Richard and Rose, Andrew, (1991), An Empirical Assessment of Non-Linearities in Models of Exchange Rate Determination, The Review of Economic Studies, 58, issue 3, p. 603-619.
16. Chinn, M. D. (1991). Some linear and nonlinear thoughts on exchange rates. Journal of International Money and Finance, 10(2), 214-230. https://doi.org/10.1016/0261-5606(91)90036-J
17. Chinn, M. D., & Meese, R. A. (1995). Banking on currency forecasts: How predictable is change in money? Journal of International Economics, 38(1–2), 161-178. https://doi.org/10.1016/0022-1996(94)01334-O
18. Chatfield, C. (1993). Calculating Interval Forecasts. *Journal of Business & Economic Statistics*, *11*(2), 121–135. https://doi.org/10.2307/1391361
19. Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, *39*(4), 841–862. https://doi.org/10.2307/2527341
20. Engel, C., & West, K. D. (2005). Exchange rates and fundamentals. Journal of Political Economy, 113(3), 485-517. https://doi.org/10.1086/429137
21. Jongen, R., Verschoor, W.F.C. and Wolff, C.C.P. (2008), FOREIGN EXCHANGE RATE EXPECTATIONS: SURVEY AND SYNTHESIS. Journal of Economic Surveys, 22: 140-165. https://doi.org/10.1111/j.1467-6419.2007.00523.x