1 MORE PREDICTABLE THAN EVER, WITH THE WORST MSPE EVER[1]

Pablo Pincheira-BROWN [2]

Nicolás HARDY [3]

Abstract

A fairly common approach to evaluate if a given time series Y_{t+1} is predictable, compares the Mean Squared Prediction Error (MSPE) of a plausible predictor for Y_{t+1} and the MSPE of a naïve benchmark like a constant forecast or the historical average of the predictand, which display zero or a small covariance with the target variable. If the MSPE of the plausible predictor is lower than that of the benchmark, Y_{t+1} is considered predictable, otherwise is considered unpredictable. This intuitive and standard approach might not be truly capturing the essence of predictability, which in words of some authors refers to a notion of dependence between the target variable and variables or events that happened in the past. In particular, when the plausible forecast under evaluation is inefficient, it might face a paradoxical situation: On the one hand, it could have a strong and positive correlation with the target variable, much greater than the correlation of the benchmark with the same target variable. Yet, on the other hand, it could be outperformed in terms of MSPE by the same naïve benchmark. We propose to evaluate predictability directly, with a simple test based on the covariance between the forecast and the target variable. Using Monte Carlo simulations we study size and power of three variations of this test. In general terms, they all behave reasonably well. We also compare their behavior with a traditional test of equality in MSPE. We show that our covariance tests can detect predictability even when MSPE comparisons do not. Finally, we illustrate the relevance of our observation when forecasting monthly oil returns with a forecast based on the Chilean peso.

Keyword: Mean Squared Prediction Error, Correlation, Forecasting, Time Series, Random Walk

JEL Classification: C52, C53, G17, E270, E370, F370, L740, O180, R310

1. Introduction

A fairly common approach to evaluate if a given time series Y_{t+1} is predictable, compares the Mean Squared Prediction Error (MSPE) of a plausible predictor for Y_{t+1} and the MSPE of a naïve benchmark displaying zero or a small covariance with the target variable. In other words, a benchmark consistent with the notion of unpredictability for Y_{t+1}[4]. Typically the naïve benchmark

---

[1] *Corresponding Author: Pablo Pincheira-Brown: Diagonal Las Torres 2640, Peñalolén, Santiago, Chile. Email: pablo.pincheira@uai.cl.*

[2] *School of Business Universidad Adolfo Ibáñez.*

[3] *Facultad de Administración y Economía Universidad Diego Portales.*

[4] *An AR(p) for instance, would not be a benchmark consistent with unpredictability, as this process is indeed predictable. As mentioned in the article, typical naïve benchmarks consistent with the idea of unpredictability of a stationary series are a constant forecast or the historical mean of the predictand.*

is a constant forecast or some sort of historical average of the predictand. If the MSPE of the plausible predictor is lower than that of the benchmark, the series is said to be predictable, otherwise, it is said to be unpredictable (at least relative to the information set used to build the plausible forecast). One problem with this intuitive and standard approach is that it might not be truly capturing the essence of predictability, which in the words of some authors, refers to a notion of dependence between the predictand and variables or events that happened in the past.

For instance, Diebold and Kilian (2001) mention: "*The extent of a series´ predictability depends on how much information the past conveys regarding future values of this series*;" Diebold and Kilian (2001) page 657. While the same authors mention that "*It is natural and informative to judge forecasts by their accuracy*" Diebold and Kilian (2001) page 657, accuracy and predictability are different, albeit, related concepts. The first quote of Diebold and Kilian (2001) claims that predictability is implicitly defined as a connection between past information and future values of a series. Similarly, Clements and Hendry (1998) make use of the definition of statistical independence of a random variable, relative to an information set, to define their notion of unpredictability. If the conditional and unconditional distribution of the target variable are the same, then this variable is considered unpredictable, at least with respect to the available information set. The authors also generalized this definition to the case of mean unpredictability and variance unpredictability, but the basic notion is the same.

So, both Diebold and Kilian (2001) and Clements and Hendry (1998) base their core notions of predictability on the connection between past and future and not on a measure of accuracy per se. In particular, when the plausible forecast under evaluation is inefficient, it might face a paradoxical situation: On the one hand, it could have a strong and positive correlation with the target variable, much greater than the correlation of the benchmark with the same target variable. Yet, on the other hand, it could be outperformed in terms of MSPE by the same naïve benchmark.

In this paper, we propose to evaluate predictability in the simplest possible way: with a test based on the covariance between the forecast and the target variable[5]. We study size and power of three versions of such a test via Monte Carlo simulations. In general terms, they all behave reasonably well. We also compare their behavior with traditional tests of equality in MSPE and show that they can provide very different results. To be more precise: when the plausible forecast is Mincer and Zarnowitz (1969) efficient, it is expected that a positive covariance with the target variable will also lead to a lower MSPE relative to that of the naïve population mean forecast. Nevertheless, as shown by Pincheira and Hardy (2024a), if the plausible forecast is not efficient, a positive covariance with the target variable, which clearly indicates predictability, does not necessarily imply a lower MSPE relative to a naïve benchmark.

We illustrate the relevance of our observation with an out-of-sample exercise in which monthly oil returns are predicted with semiannual returns of a commodity-currency. Despite the strong correlation with future oil returns, the MSPE associated to this interesting forecast is higher to that of naïve benchmarks which in turn display little or no covariance at all with the target variable. The same situation happens when forecasting monthly returns of propane and heating oil.

The rest of this paper is organized as follows. In section 2 we show simple examples in which forecasts, strongly correlated with the target variable, do not fare well in terms of MSPE compared to naïve counterparts with no dependence whatsoever with the predictand. In Section 3 we present the tests that we will use to evaluate the covariance between the target value and the forecast itself. We also evaluate their size and power with Monte Carlo simulations. In section 4 we illustrate these paradoxical results with empirical applications. Finally, section 5 concludes.

[5] *Our focus on developing a dependency test rather than a test based on forecast accuracy, is similar in spirit to the approach followed by Pesaran and Timmermann (1992, 2009), although, of course, in a somewhat different context.*

2. Predictability vs Forecast Accuracy

2.1 Simple examples

In this section we illustrate with simple examples that predictability might not be associated to reductions in MSPE relative to a naïve and independent benchmark. In all these three examples our candidate forecasts share a common feature: Mincer and Zarnowitz (1969) inefficiency. Our discussion here follows the line of "The MSPE Paradox" depicted in Pincheira and Hardy (2024a, 2024b).

Let Y_t be a mean zero target variable with positive variance. At time t, we have two competing forecasts X_{t-1} and Z_{t-1} for Y_t. It is important to notice that both X_{t-1} and Z_{t-1} are variables constructed with information previous to time *t* and that they are taken as primitives (hence, we are not concerned here about parameter uncertainty). Let us also assume that the vector $(Y_t, X_{t-1}, Z_{t-1})'$ has finite second moments and is weakly stationary. For clarity of exposition, we will drop the sub-indexes *t and t-1* hereafter.

<u>Example 1</u>: Let us consider the case in which $E(Y) = E(X) = 0$ and Z is a zero-forecast. Consequently $V(Z) = Cov(Y, Z) = E(Z| \quad |2) = 0$. Furthermore, let us also assume that V(X) = V(Y) > 0 and that $Cov(Y, X) = 0.5V(Y) > 0$ so that we have $Corr(Y, X) = 0.5$. It is straightforward to see that the covariance matrix of the (Y,X) vector is positive-definite and that there is a strong connection between forecast X and the target variable Y, so that Y is predictable. Nevertheless, both forecasts Z and X will have the same MSPE:

$$\Delta MSPE \equiv MSPE_X - MSPE_Z = E(Y - X)^2 - E(Y - Z)^2$$

$$\Delta MSPE = V(Y - X) - V(Y) = V(Y) + V(X) - 2Cov(Y, X) - V(Y) = 0$$

So, by looking at MSPE differentials only, we would conclude that forecast X is not useful to predict Y, or even worse, if for some reason X is expected to be the best forecast for Y, we could take one step further and conclude that Y is unpredictable. Yet the strong correlation between X and Y says otherwise.

The key problem here is that X is inefficient as its forecast error is also correlated with X:

$$Cov(Y - X, X) = Cov(Y, X) - V(X) = 0.5V(Y) - V(Y) = -0.5V(Y) < 0$$

The use of MSPE in this case, indicates that X and the zero forecast are predictors that on average fare equally well in terms of accuracy, despite the fact that the zero forecast is totally independent of the target variable, while X has a strong connection with it. Unfortunately, in this example, MSPE comparisons would fail to detect the usefulness of forecast X, due to its sizable correlation with its own forecast error.

<u>Example 2:</u> Consider the same example 1 with a minor variation: now $Cov(Y, X) = 0.25V(Y) > 0$. Then we will have a more extreme situation than before, as the useful forecast X will be outperformed in terms of MSPE by the naïve zero forecast:

$$\Delta MSPE \equiv MSPE_X - MSPE_Z = E(Y - X)^2 - E(Y - Z)^2$$

$$V(Y - X) - V(Y) = V(Y) + V(X) - 2Cov(Y, X) - V(Y) = 0.5V(Y) > 0$$

So, by looking at MSPE differentials only, we would conclude that forecast X is no useful to predict Y, and as before, we could even conclude that Y is unpredictable. Yet, the strong correlation between X and Y again proves that this conclusion would be incorrect.

<u>Example 3:</u> Let us now consider a case in which $E(Y) = E(X) > 0$ and Z is a zero-forecast. As usual $V(Z) = Cov(Y, Z) = E(Z| \quad |2) = 0$, but differently from the previous two examples, now the zero forecast Z is biased. Furthermore, let us also assume that 0 < V(X) < $min\{[E(Y)]^2, V(Y)\}$, and that Y and X are independent random variables. This implies $Cov(Y, X) = 0$. In this case it is clear that there is no connection between forecast X and the target variable Y so that Y is unpredictable by any measurable function of X. Nevertheless this independent variable X will outperform the zero forecast in terms of MSPE:

$$\Delta MSPE \equiv MSPE_X - MSPE_Z = E(Y-X)^2 - E(Y-Z)^2$$
$$V(Y-X) - E[Y^2] = V(Y) + V(X) - 2Cov(Y,X) - E[Y^2]$$
$$V(Y) + V(X) - E[Y^2] = V(X) - [E(Y)]^2 < 0$$

So, by looking at MSPE differentials only, we would conclude that X outperforms the naïve zero forecast and from that point of view we could say that Y is predictable[6]. Yet, X and Y are independent random variables that have no connection whatsoever. Here the comparison in terms of MSPE is correct but it is only driven by bias. As both forecasts are independent with Y, the only reason why X outperforms Z is because it has a lower bias. Y is unpredictable by X, yet this usual comparison of MSPE differentials is unable to detect this important result.

2.2 A slightly more general case

While examples 1-3 clearly illustrate that predictability might not be strictly related to forecast accuracy in terms of MSPE, we are particularly interested in cases in which a forecast X, positively correlated with Y, is less accurate than the naïve population mean forecast for Y. So, let us now show a slightly more general picture of the problem. We will analyze the case in which Y and X have positive variances but Z is just a constant c=E(Y). Here, we will have that Z is an unbiased and efficient forecast for Y.

It is straightforward to show that:

$$MSPE_X \equiv E(Y-X)^2 = V(Y-X) + [E(Y-X)]^2 = V(Y) + V(X) - 2Cov(Y,X) + [E(Y-X)]^2$$
$$MSPE_Z \equiv E(Y-Z)^2 = E\big(Y-E(Y)\big)^2 = V(Y)$$

Suppose $\Delta MSPE \equiv MSPE_X - MSPE_Z$= 0, so that by simply looking at MSPE differentials we would conclude that Y is not predictable by X. In this case we would have that

$$0 = MSPE_X - MSPE_Z = V(Y) + V(X) - 2Cov(Y,X) + [E(Y-X)]^2 - V(Y)$$
$$0 = \Delta MSPE = V(X) - 2Cov(Y,X) + [E(Y-X)]^2$$

Or

$$Cov(Y,X) = \frac{V(X)+[E(Y-X)]^2}{2} > 0 \quad (1)$$

So equality in MSPE between forecast X and the naïve forecast Z is perfectly compatible with a positive covariance between X and the predictand.

Again, this is possible as long as X is inefficient. Otherwise we would have $Cov(Y,X) = V(X)$ and $E(X) = E(Y)$ and the MSPE of forecast X would have been lower than V(Y):

$$\Delta MSPE \equiv MSPE_X - MSPE_Z = V(Y) - V(X) - V(Y) = -V(X) < 0$$

This again indicates that predictability might be at odds with forecast accuracy when one of the forecasts in the competition is inefficient.

By looking at these very simple derivations it is clear that if a researcher is interested in detection of predictability, he/she should look beyond forecast accuracy unless he/she could make sure that all the forecasts under analysis satisfy Mincer-Zarnowitz efficiency. Yet, from our literature review, we have observed that violations of efficiency are fairly common[7]. Furthermore, it is also

---

[6] *This would be the case if the researcher believes that the unconditional expectation of Y is zero.*

[7] *Deviations from efficiency are found across multiple variables in a number of articles. See for instance, Ince and Molodtsova (2017); Joutz and Stekler (2000); Ang, Bekaert and Wei (2007); Patton and Timmermann (2012); Nordhaus (1987), Pincheira, Bentancor and Hardy (2023) and Pincheira (2012, 2010) just to mention a few.*

frequent to find papers do not even mentioning analyses of forecast efficiency. In light of this evidence, we think that it is relevant to explore predictability directly, by measuring the degree of dependence between the forecasts and the target variable. In the next section, we analyze size and power of three simple tests evaluating the covariance between a target variable and a forecast.

3. Tests and Monte Carlo Simulations

3.1 Our null hypothesis

We focus on the following hypotheses:

$$H_0: Cov(Y, X) = 0 \quad (2)$$
$$H_A: Cov(Y, X) > 0 \quad (3)$$

Our null hypothesis posits no linear relationship between the forecast and the target variable. No rejection of the null is consistent with the absence of predictability for Y, at least with information based on X. On the contrary, rejection of the null is a clear indication of predictability, as the forecast X is linearly related to the target variable.

A few observations are worth mentioning: First, we focus on one-sided tests because we expect a reasonable forecast to be positively related to the target variable. In case of a negative and significant covariance, the researcher may consider the obvious multiplication of his/her favorite forecast by minus one. Second, as the covariance of a random variable with a constant is exactly zero, our null hypothesis could be interpreted as a comparison between the covariance of the target variable with forecast X and the covariance of the target variable with a constant forecast. This includes the traditional zero forecast or the unknown population mean of the target variable. Under quadratic loss, the population mean is the optimal long-term forecast of the stationary target variable Y. So we indeed will be capturing the idea of Diebold and Kilian (2001), who propose to evaluate the relative accuracy of short versus long-term forecasts. Besides, a constant forecast will be optimal even in the short-run for a martingale in difference process, which is widely used in economics and finance. Third, our approach is also consistent with the definition of unpredictability proposed by Clements and Hendry (1998). In their definition, mean unpredictability relative to the information contained in X is expressed in terms of the equality

$$E(Y|X) = E(Y)$$

This implies that $Cov(Y, X) = Cov\big(Y, E(Y|X)\big) = Cov\big(Y, E(Y)\big) = 0$. And we can apply the same argument for their definition of variance unpredictability.

Fourth, if forecast X is Mincer-Zarnowitz efficient, then rejection of our null hypothesis in (2) also implies that forecast X outperforms the population mean forecast, the zero forecast and any other possible constant forecast under quadratic loss. If our null hypothesis (2) is correct, however, and we assume that forecast X has positive variance, the only direct implications are that forecast X is outperformed in terms of MSPE by the population mean forecast and that X cannot be Mincer-Zarnowitz efficient. To see the first point notice that

$$MSPE_X = E[Y - X]^2 = V(Y - X) + (E[Y - X])^2 = V(Y) + V(X) + (E[Y - X])^2 > V(Y) = MSPE_Z$$

Notice also that Mincer Zarnowitz efficiency implies $Cov(Y, X) = V(X)$, which, under the null of a zero covariance between X and Y, is an open contradiction with the positive variance assumption for X.

3.2 Our tests

We use three different statistics to test our null hypothesis. The first two are obtained in regression form. They require the following assumptions:

A1) Strong stationarity for the pair (Y_t,X_t).

A2) Finite second moments for the vector (Y_t,X_t).

A3) $E|Y_t|^r < \infty$, $E|X_t|^r < \infty$, for some $r > 4$.

A4) The mixing coefficients $\alpha(l)$ for the pair (Y_t,X_t) are such that $\sum_{l=1}^{\infty} \alpha(l)^{1-\frac{4}{r}} < \infty$.

These assumptions are necessary to obtain asymptotic normality of the OLS estimates in the simple regression frameworks that we use next. See Hansen (2022), page 501.

1) We consider the following regression:

$$Y = a + bX + e, E(e) = E(eX) = 0$$

which implies $b = \frac{Cov(Y,X)}{V(X)}$. Under assumptions A1)-A4), we know that the simple t-statistic associated to the *b* parameter, computed with HAC standard errors, will follow a standard normal distribution, so we use this simple t-statistic as a first approach[8]. We will refer to this test simply as “regression test”.

2) We also consider the inverse regression:

$$X = \alpha + \beta Y + \varepsilon, E(\varepsilon) = E(\varepsilon X) = 0$$

Under assumptions A1)-A4), we know that the simple t-statistic associated to the β parameter, computed with HAC standard errors, will follow a standard normal distribution, so we use this simple t-statistic as a second approach using HAC standard errors according to Newey and West (1987, 1994). Notice also that this β parameter could be expressed as

$$\beta = \frac{Cov(Y,X)}{V(Y)}$$

When X is Mincer-Zarnowitz efficient, we will have that $Cov(Y,X) = V(X)$, so in this case β can have the interpretation of a coefficient of determination, as under efficiency of X, β represents the share of V(Y) “explained” by forecast X, which is a textbook definition of a coefficient of determination. Consequently, we will refer this test as a R^2 test.

3) We consider the following statistic:

$$Correlation - test = \sqrt{T}\frac{r}{\sqrt{\Delta h' \Delta g' \sum_{j=-\infty}^{\infty} \Omega_j \, \Delta g \Delta h}}$$

where $\Omega_j =$

$$\begin{bmatrix} Cov(X_t, X_{t-j}) & Cov(Y_t, X_{t-j}) & Cov(X_t^2, X_{t-j}) & Cov(Y_t^2, X_{t-j}) & Cov(X_tY_t, X_{t-j}) \\ Cov(X_t, Y_{t-j}) & Cov(Y_t, Y_{t-j}) & Cov(X_t^2, Y_{t-j}) & Cov(Y_t^2, Y_{t-j}) & Cov(X_tY_t, Y_{t-j}) \\ Cov(X_t, X_{t-j}^2) & Cov(Y_t, X_{t-j}^2) & Cov(X_t^2, X_{t-j}^2) & Cov(Y_t^2, X_{t-j}^2) & Cov(X_tY_t, X_{t-j}^2) \\ Cov(X_t, Y_{t-j}^2) & Cov(Y_t, Y_{t-j}^2) & Cov(X_t^2, Y_{t-j}^2) & Cov(Y_t^2, Y_{t-j}^2) & Cov(X_tY_t, Y_{t-j}^2) \\ Cov(X_t, Y_{t-j}X_{t-j}) & Cov(Y_t, Y_{t-j}X_{t-j}) & Cov(X_t^2, Y_{t-j}X_{t-j}) & Cov(Y_t^2, Y_{t-j}X_{t-j}) & Cov(X_tY_t, Y_{t-j}X_{t-j}) \end{bmatrix}$$

[8] *HAC standard errors are computed according to Newey and West (1987, 1994).*

$$\Delta g = \begin{bmatrix} -2m_X & 0 & -m_Y \\ 0 & -2m_Y & -m_X \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\Delta h = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{S_Y S_X} \end{bmatrix}$$

T represents the total number of observations, r denotes the sample correlation between X and Y and m_Y and m_X are the sample means of the target variable Y and the forecast X respectively. Finally, S_X and S_Y are the sample standard deviations of X and Y respectively. To estimate the long run variance $\sum_{j=-\infty}^{\infty} \Omega_j$ we use the traditional consistent estimator of Newey and West (1987).

Using the multivariate Delta method and mild assumptions, it is possible to show that under the null hypothesis of a zero correlation between forecast X and the target variable Y, the Correlation test is asymptotically standard normal. The formal derivation of the test is in Appendix A1. We will simply label this test as "Correlation test".

Notice that this test could be written equivalently in a simpler way, provided that S_X and S_Y are consistent estimators of the standard deviations of X and Y respectively.

$$Correlation - test = \frac{\sqrt{T}\, S_{XY}/(S_X S_Y)}{(S_X S_Y)^{-1}\sqrt{\nabla g' \sum_{j=-\infty}^{\infty} \Omega_j \nabla g}}$$

where $\Omega_j = \begin{bmatrix} Cov(X_t, X_{t-j}) & Cov(Y_t, X_{t-j}) & Cov(X_t Y_t, X_{t-j}) \\ Cov(X_t, Y_{t-j}) & Cov(Y_t, Y_{t-j}) & Cov(X_t Y_t, Y_{t-j}) \\ Cov(X_t, Y_{t-j} X_{t-j}) & Cov(Y_t, X_{t-j} Y_{t-j}) & Cov(X_t Y_t, X_{t-j} Y_{t-j}) \end{bmatrix}$; $\nabla g \begin{bmatrix} m_X \\ m_Y \\ m_{YX} \end{bmatrix} = \begin{bmatrix} -m_Y \\ -m_X \\ 1 \end{bmatrix}$

Here S_{XY} is the sample covariance between the X and Y. To estimate the long run variance $\sum_{j=-\infty}^{\infty} \Omega_j$ we use the traditional consistent estimator of Newey and West (1987).

3.3 Monte Carlo Simulations

Here we explore size and power of our tests via Monte Carlo simulations. We consider two different DGPs calibrated to match linear models in which commodity returns are being predicted with either returns of a commodity-currency or with their own lags.

DGP1: In this DGP we generate data for the predictand Y_{t+1} and for our forecast X_t using a stationary and Gaussian VAR(1) process with a diagonal autoregressive matrix. For all practical matters, our DGP looks like two separate AR(1) models for each variable:

$$Y_{t+1} = \alpha_Y + \gamma_Y Y_t + \varepsilon_{t+1}^Y \qquad (4)$$

$$X_t = \alpha_X + \gamma_X X_{t-1} + \varepsilon_t^X \qquad (5)$$

Where $\begin{pmatrix} \varepsilon_{t+1}^Y \\ \varepsilon_t^X \end{pmatrix}$ is a Gaussian two dimensional with noise process with variance $V = \begin{bmatrix} \sigma_{\varepsilon^Y}^2 & \rho\sigma_{\varepsilon^Y}\sigma_{\varepsilon^X} \\ \rho\sigma_{\varepsilon^Y}\sigma_{\varepsilon^X} & \sigma_{\varepsilon^X}^2 \end{bmatrix}$, and $\rho = Corr(\varepsilon_{t+1}^Y, \varepsilon_t^X)$.

Notice that

$$Cov(Y_{t+1}; X_t) = Cov(\alpha_Y + \gamma_Y Y_t + \varepsilon_{t+1}^Y; \alpha_X + \gamma_X X_{t-1} + \varepsilon_t^X) = \gamma_Y \gamma_X Cov(Y_t; X_{t-1}) + Cov(\varepsilon_{t+1}^Y; \varepsilon_t^X)$$

Our stationary assumption implies that

$$Cov(Y_{t+1}; X_t) = \frac{\rho \sigma_{\varepsilon^Y} \sigma_{\varepsilon^X}}{1 - \gamma_Y \gamma_X}$$

Besides $MSPE = E(Y_{t+1} - X_t)^2 = Var(Y| \quad |t + 1 - X_t) +$

$$MSPE = E(Y_{t+1} - X_t)^2 = Var(Y| \quad |t + 1) + Var(X_t) - 2Cov(Y_{t+1}; X_t) +$$

$$MSPE = \frac{\sigma_{\varepsilon^Y}^2}{1 - \gamma_Y^2} + \frac{\sigma_{\varepsilon^X}^2}{1 - \gamma_X^2} - 2Cov(Y_{t+1}, X_t) + \left[\frac{\alpha_Y}{1 - \gamma_Y} - \frac{\alpha_X}{1 - \gamma_X}\right]^2$$

Under the null hypothesis of zero correlation between the forecast and the target variable we will have $MSPE = \frac{\sigma_{\varepsilon^Y}^2}{1-\gamma_Y^2} + \frac{\sigma_{\varepsilon^X}^2}{1-\gamma_X^2} + \left[\frac{\alpha_Y}{1-\gamma_Y} - \frac{\alpha_X}{1-\gamma_X}\right]^2 > \frac{\sigma_{\varepsilon^Y}^2}{1-\gamma_Y^2} \equiv MSPE_\mu$

Where $MSPE_\mu$ denotes the MSPE of predicting Y_{t+1} with its population mean.

This result is interesting, as it says that when the forecast X_t is uncorrelated with the target variable, its MSPE will be higher than when predicting Y_{t+1} with its population mean. We can also see that when the covariance between the forecast and the target variable increases, there is a good chance for the forecast X_t to outperform the population mean in terms of MSPE.

In our simulations, we calibrate the parameters of DGP1 to match a model in which Y_{t+1} represents monthly returns in WTI oil price and X_t represents the first monthly lag of semiannual Chilean Peso returns (multiplied by minus one). According to Alquist, Kilian and Vigfusson (2013) commodity–currencies have an important ability to predict oil prices. Pincheira, Bentancor, Hardy and Jarsún (2022) corroborate this finding using lags of long returns of the Chilean Peso, a commodity-currency strongly and negatively correlated to copper prices. Table 1 next reports our estimated parameters.

Table 1: Estimated parameters for DGP1

α_Y	γ_Y	α_X	γ_X	ρ	$\sigma_{\varepsilon^Y}^2$	$\sigma_{\varepsilon^X}^2$
0.280	0.147	0.205	0.835	0.130	119.858	23.136

Notes: Table 1 reports estimates of the parameters in expressions (4) and (5) in the text. To obtain these estimates we consider monthly data of WTI oil price and of the Chilean Peso from January 2000 until February 2024. Y_{t+1}in expression (4) represents monthly returns of WTI oil price, computed as the log difference between the closing oil price on the last day of month t+1, and the closing oil price on the last day of month t. X_t in expression (5) represents minus one times the semiannual log difference in the Chilean peso, computed as the log difference of the closing price of the Chilean Peso from the last day in month t, and the closing price of the Chilean peso from the last day of month t-6. All our data is obtained at a daily frequency from Refinitiv Datastream.

On the one hand, from our data we get an estimate of $\rho = 0.13$ which is consistent with a positive correlation of roughly 8% between our forecast and the predictand. Therefore, in terms of our main null and alternative hypotheses (2) and (3) our estimates indicate a situation which is more likely to be consistent with the alternative hypothesis in (3).

On the other hand, our same estimates from Table 1 imply a RMSPE (Root Mean Squared Prediction Error) of roughly 13.6 for forecast X_t, which is higher than the implied standard deviation of the target variable of 11.07, which coincides with the RMSPE of predicting with the population mean. So, from a traditional point of view of testing equality in MSPE, our estimated parameters suggest that a null of equality in MSPE would be rejected in favor of the population mean forecast. In other words our computation of correlations and MSPE go in opposite directions, as in the MSPE Paradox discussed by Pincheira and Hardy (2024a).

To explore size and power of our tests we consider different values of the ρ parameter. For size, we set $\rho = 0$, which implies a zero correlation between forecast X_t and Y_{t+1}. For experiments evaluating power we consider positive values for the ρ parameter, which in turns translates into positive correlations between X_t and Y_{t+1}. While the main focus in our experiments is to evaluate the behavior of our three correlation tests, we also keep track of MSPE to evaluate the null of equality in MSPE associated to forecast X_t and the population mean forecast. This allows us to compare the behavior of these two families of tests when trying to evaluate predictability: correlation-based tests and MSPE-based tests.

We consider 5000 replications in each Monte Carlo simulation. Tables 2 and 3 show our results for two different sample sizes: T=200 and T=1000. Notice that this is a pure out-of-sample analysis, as we are generating both the target variable Y_{t+1} and the forecast X_t. Therefore there is no need to estimate parameters to generate our forecasts. This is a scenario in which forecasts are available naturally, as in the case of forecast surveys, financial variables that serve as leading indicators or private or public forecasts from international of domestic institutions, as Central Banks and the IMF. We focus on one-step-ahead forecasts only leaving the analysis of multi-step-ahead forecasts as an extension for further research.

Table 2: Size and Power of our Correlation Tests when T=200 and nominal size is 10%

$\rho = Corr(\varepsilon^Y_{t+1}, \varepsilon^X_t)$	**0 (Size)**	**0.13**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
Theoretical $Corr(Y_{t+1}, X_t)$	0.000	0.081	0.124	0.186	0.248	0.310	0.372	0.434	0.496	0.558
One sided regression test	0.118	0.419	0.649	0.879	0.981	1.000	1.000	1.000	1.000	1.000
One sided R^2 test	0.114	0.421	0.650	0.886	0.982	1.000	1.000	1.000	1.000	1.000
One sided Correlation test	0.112	0.420	0.651	0.886	0.982	1.000	1.000	1.000	1.000	1.000

Notes: The first row in Table 2 reports the theoretical correlation between forecast X_t and Y_{t+1}. The rest of the rows present rejection rates of the null hypothesis of zero covariance between forecast X_t and Y_{t+1}. These rejection rates are obtained across all our 5000 replications for a nominal size of 10%, when the sample size of our available data is 200. The first numerical column in Table 2 displays the empirical size of each test in the last three rows of the table.

Table 3: Size and Power of our Correlation Tests when T=1000 and nominal size is 10%

$\rho = Corr(\varepsilon^Y_{t+1}, \varepsilon^X_t)$	**0 (Size)**	**0.13**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
Theoretical $Corr(Y_{t+1}, X_t)$	0.000	0.081	0.124	0.186	0.248	0.310	0.372	0.434	0.496	0.558
One sided regression test	0.103	0.838	0.987	1.000	1.000	1.000	1.000	1.000	1.000	1.000
One sided R^2 test	0.101	0.839	0.987	1.000	1.000	1.000	1.000	1.000	1.000	1.000
One sided Correlation test	0.101	0.838	0.987	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: The first row in Table 3 reports the theoretical correlation between forecast X_t and Y_{t+1}. The rest of the rows present rejection rates of the null hypothesis of zero covariance between forecast X_t and Y_{t+1}. These rejection rates are obtained across all our 5000 replications for a nominal size of 10%, when the sample size of our available data is 1000. The first numerical column in Table 3 displays the empirical size of each test in the last three rows of the table.

The first row (in bold) in Tables 2 and 3 displays the different values of the ρ parameter that we consider in our simulations[9]. In the second row in the tables we see how the different values of the ρ parameter generate different correlations between the target variable and our forecast X_t. In the rest of the rows in both tables we show rejection rates of the null hypothesis obtained for each test in our 5000 replications using a 10% significance level. When $\rho = 0$, the correlation between our forecast and the target variable is exactly zero, so our null hypothesis (2) holds true. Consequently, rejection rates correspond to the size of the tests. In Table 2 we see that all three correlation tests are slightly oversized when T=200, as rejection rates are in the range between 11% and 12%. As the sample size gets larger, Table 3 reveals that all three tests are very close to nominal size. Interestingly, the best behavior in terms of size is achieved by the correlation test derived by the Delta method.

The rest of the columns in Tables 2 and 3 show the power of the tests for different values of the correlation between the forecast and the target variable. As expected, with a larger sample size, all three tests display more power. It is interesting to note that even when T=200, all three tests are relatively powerful as it only takes a correlation of 0.31 between the forecast and the target variable to get a perfect rejection rate of the null hypothesis. When T=1000, Table 3 reveals that it only takes a correlation of 0.186 to get full power. All in all our three tests seem to be powerful and correctly sized.

Tables 4 and 5 show the RMSPE of our forecasts, the correlation of forecast X_t with the target variable, rejection rates of our Correlation test computed with the Delta method and the traditional Diebold and Mariano (1995) and West (1996) test (henceforth DMW) of inequality in MSPE[10]. Here we use as a benchmark forecast the population mean of the target variable $\mu = \frac{\alpha_Y}{1-\gamma_Y}$ which is just a constant with zero covariance with the target variable. To be more precise, while the main focus of Tables 2 and 3 is on the power and size of our three correlation tests, in Tables 4 and 5 we also consider the following null and alternative hypotheses of inequality in MSPE:

H1: $MSPE_\mu \leq MSPE_X$
H2: $MSPE_\mu > MSPE_X$

Where $MSPE_\mu$ represents the MSPE resulting from the prediction of Y_{t+1} with its population mean, and $MSPE_X$ represents the MSPE resulting from the prediction of Y_{t+1} with X_t. Notice that for clarity of exposition, in Tables 4 and 5 we only include the correlation test derived with the Delta method[11].

Table 4: MSPE based test vs Correlation based test when T=200 and nominal size is 10%.

$\rho = Corr(\varepsilon_{t+1}^Y, \varepsilon_t^X)$	0.00	0.13	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean RMSPE benchmark	11.055	11.055	11.055	11.055	11.055	11.055	11.055	11.055	11.055	11.055
Mean RMSPE of Forecast X_t	14.119	13.554	13.240	12.778	12.298	11.799	11.277	10.732	10.158	9.551

[9] *Tables 2-3 show results for several values of $\rho = Corr(\varepsilon_{t+1}^Y, \varepsilon_t^X)$, including $\rho = 0.13$, which is the estimate obtained from our data.*

[10] *We carried out the DMW test in the spirit of the article by Giacomini and White (2006).*

[11] *Those readers familiar with the nested model literature might be tempted to think that it would be more appropriate to consider tests like those of Clark and West (2006, 2007) or Clark and McCracken (2001). Yet, as we are not using models to build our forecasts, that is not necessary, so we simply compare MSPE between two forecasts with a DMW test using HAC standard errors according to Newey and West (1987, 1994).*

Theoretical $Corr(Y_{t+1}, X_t)$	0.000	0.081	0.124	0.186	0.248	0.310	0.372	0.434	0.496	0.558
One sided DMW test	0.000	0.000	0.000	0.000	0.002	0.016	0.086	0.279	0.641	0.949
One sided Correlation test	**0.112**	**0.420**	**0.651**	**0.886**	**0.982**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

Notes: Similar to Tables 2 and 3, the top row in Table 4 displays the different correlations $Corr(\varepsilon_{t+1}^{Y}, \varepsilon_t^{X})$ that we use in our simulations. The rows under the labels "Mean RMSPE benchmark" and "Mean RMSPE of Forecast X_t" represents the average RMSPE obtained across our 5000 replications for the population mean benchmark and forecast X_t respectively. The row under the label "Theoretical $Corr(Y_{t+1}, X_t)$" presents the theoretical correlation between forecast X_t and the target variable. This row is exactly the same second row in Tables 2 and 3. The row under the label "One sided DMW test" presents rejection rates of the null on inequality in MSPE between the benchmark and forecast X_t. Rejection rates favor precisely this latter forecast X_t. The last row in Table 4 is just the same last row in Table 2. In all these exercises the nominal size is 10%, the number of replications is 5000 and the length of our data is T=200.

Table 5: MSPE based test vs Correlation based test when T=1000 and nominal size is 10%.

$\rho = Corr(\varepsilon_{t+1}^{Y}, \varepsilon_t^{X})$	0.00	0.13	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean RMSPE benchmark	11.067	11.067	11.067	11.067	11.067	11.067	11.067	11.067	11.067	11.067
Mean RMSPE of Forecast X_t	14.133	13.570	13.256	12.795	12.317	11.819	11.299	10.754	10.180	9.572
Theoretical $Corr(Y_{t+1}, X_t)$	0.000	0.081	0.124	0.186	0.248	0.310	0.372	0.434	0.496	0.558
One sided DMW test	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.469	0.984	1.000
One sided Correlation test	**0.101**	**0.838**	**0.987**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

Notes: This table presents the same information provided in Table 4. The only difference is that Table 5 reports simulation results when our sample size is bigger, T=1000.

From Tables 4 and 5 it is striking to see that MSPE differentials and correlations go in opposite directions. For instance, the null hypothesis of zero correlation between the forecast and the target variable holds true even though the RMSPE of the benchmark forecast is much lower than that of the X_t forecast $(11.055 vs 14.119 \in Table 4 \wedge 11.067 vs 14.133 \in Table 5)$. In both Tables we show that out of the 5000 replications the DMW test cannot reject the null even once in favor of the forecast X_t even though both predictors: the constant population mean and X_t have zero covariance with the target variable.

The rest of the columns in Tables 4 and 5 are even more interesting because they show cases with a positive correlation between the forecast X_t and the target variable. In particular, whenever $\rho > 0$we see strong and positive correlations between X_t and Y_{t+1}. Despite these important positive correlations we still observe a lower MSPE for the constant benchmark model whenever $0 < \rho < 0.7$. Put differently, in all this range of values for ρ, the constant benchmark model is more accurate than X_t in terms of MSPE, yet Y_{t+1} is clearly predictable by X_t, as they are strongly connected with relatively high correlations. This is another example in which MSPE comparisons fail to detect predictability. Here there is no one to blame. DMW does its duty and tends to correctly reject in favor of the forecast with a lower MSPE. Similarly, our correlation test does its duty also, rejecting in favor of the forecast with a higher correlation. The problem again is that our forecast X_t is Mincer-Zarnowitz inefficient. In this scenario, traditional MSPE comparisons may fail to detect predictability as we saw earlier in our basic examples in section 2.

DGP2: In this DGP we generate data for the predictand Y_{t+1} using a simple stationary AR(1) process:

$$Y_{t+1} = \alpha_Y + \gamma_Y Y_t + \sigma u_{t+1} \qquad (6)$$

Our main forecast X_t is simply defined as the inefficient first lag of Y_{t+1}. In other words: $X_t \equiv Y_t$.

As in DGP1, here we calibrate the parameters of our simple model to match features of monthly returns of WTI oil. An important difference with DGP1 is that now our innovations u_{t+1} are not Gaussian. They follow a t-student distribution with 4 degrees of freedom, so to give higher probability to extreme outcomes. Table 6 next reports the parameters used in DGP2.

Table 6: Estimated parameters for DGP2

α_Y	γ_Y	σ	$degrees\ of\ freedom$
0.282	0.147	7.79	4

Notes: Table 6 reports estimates of the parameters in expression (6) in the text. To obtain these estimates we consider monthly data of WTI oil price from January 2000 until February 2024. Y_{t+1} in expression (6) represents monthly returns of WTI oil price, computed as the log difference between the closing oil price in the last day of month t+1, and the closing oil price in the last day of month t. All our data is obtained at a daily frequency from Refinitive Datastream.

Just like in our exercises with DGP1, here we also consider 5000 replications in each Monte Carlo simulation. Tables 7 and 8 are akin to Tables 2 and 3 and show our results for two different sample sizes: T=200 and T=1000 when the nominal size of the tests is 10%. We focus on one-step-ahead forecast only, leaving the analysis of multi-step-ahead forecasts as an extension for further research.

Table 7: Size and Power of our Correlation Tests when T=200. Nominal size is 10%. DGP2

γ_Y	0 (Size)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean $Corr(Y_{t+1}, X_t)$	-0.010	0.087	0.185	0.282	0.379	0.477	0.574	0.670	0.766	0.861
One sided regression test	0.115	0.420	0.790	0.954	0.993	0.999	1.000	1.000	1.000	1.000
One sided R^2 test	0.118	0.415	0.783	0.952	0.993	0.999	1.000	1.000	1.000	1.000
One sided Correlation test	0.107	0.401	0.767	0.952	0.992	0.999	0.999	1.000	1.000	1.000

Notes: The last three rows in Table 7 show rejection rates for our three correlation tests and for different values of the parameter γ_Y in DGP 2. These rejection rates are obtained across all our 5000 replications for a nominal size of 10%, when the sample size of our available data is T=200. The first numerical column in Table 7 displays the empirical size of the tests. The first row in Table 7 reports the average across all our 5000 replications of the correlation between our forecast X_t and the target variable.

Table 8: Size and Power of our Correlation Tests when T=1000. Nominal size is 10%. DGP2

γ_Y	0 (Size)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean $Corr(Y_{t+1}, X_t)$	-0.002	0.098	0.198	0.298	0.397	0.497	0.597	0.696	0.796	0.896
One sided regression test	0.106	0.951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
One sided R^2 test	0.106	0.951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
One sided Correlation test	0.103	0.951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: This table presents the same information provided in Table 7. The only difference is that Table 8 reports simulation results when our sample size is bigger, T=1000.

The first row in Tables 7 and 8 displays the different values of the γ_Y parameter that we consider in our simulations. In our simple AR(1) model, this parameter coincides with the correlation between our forecast X_t and the target variable. In the second row in the tables we report the average correlation $Corr(Y_{t+1}, X_t)$ across our 5000 replications. We see that this statistic is pretty close to the actual population correlation γ_Y, especially in Table 8 when T=1000.

In the rest of the rows in both tables we show the rejection rates of the null hypothesis using a 10% significance level. When γ_Y= 0 the correlation between our forecast and the target variable is exactly zero, so our null hypothesis (2) holds true. Consequently, rejection rates correspond to the size of the tests. In Table 7 we see that all three correlation tests are mildly oversized when T=200, as rejection rates are in the range between 10% and 12%. As the sample size gets larger, Table 8 reveals that all three tests are very close to nominal size. Just as in the case of DGP1, the best behavior in terms of size is achieved again by the correlation test derived by the Delta method.

The rest of the columns in Tables 7 and 8 show the power of the tests for different values $\gamma_Y > 0$. As expected, with a larger sample size, all three tests display more power. Again, even when T=200, all three tests are relatively powerful as it only takes a correlation of 0.4 between the forecast and the target variable to get an almost perfect rejection rate of the null hypothesis. When T=1000, Table 8 reveals that it only takes a correlation of 0.2 to get full power. This is important because we are using a DGP with fat tails. Tables 7 and 8 show that this is not a serious obstacle for the correlation tests as they seem to be correctly sized and relatively powerful, even with moderate sample sizes of T=200.

Tables 9 and 10 show the RMSPE of our forecasts, the mean correlation of forecast X_t with the target variable across our 5000 replications, rejection rates of our Correlation test computed with the Delta method and the traditional DMW test of inequality in MSPE (H1 vs H2). As in the case with DGP1 we use as a benchmark forecast the population mean of the target variable$\mu = \frac{\alpha_Y}{1-\gamma_Y}$. Notice that for clarity of exposition in Tables 9 and 10 we only include the correlation test derived with the Delta method.

Tables 9 and 10 show clearly that the MSPE Paradox emerges whenever $0 < \gamma_Y \leq 0.5$. For instance, the null hypothesis of zero correlation between the forecast and the target variable holds true even though the RMSPE of the benchmark forecast is much lower than that of the X_t forecast $(10.858 vs 15.342 \in Table 9 \wedge 10.978 vs 15.527 \in Table 10)$. In both Tables we show that out of the 5000 replications the DMW test cannot reject the null even once in favor of the forecast X_t even though both predictors: the constant population mean and X_t have zero covariance with the target variable.

Table 9: MSPE based test vs Correlation based test when T=200, nominal size is 10%. DGP2

γ_Y	0.00	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean RMSPE benchmark	10.858	10.911	11.079	11.377	11.838	12.522	13.542	15.143	17.956	24.458
Mean RMSPE of Forecast X_t	15.342	14.629	14.007	13.459	12.970	12.532	12.135	11.774	11.444	11.141
Mean $Corr(Y_{t+1}, X_t)$	-0.010	0.087	0.185	0.282	0.379	0.477	0.574	0.670	0.766	0.861
One sided DMW test	0.000	0.000	0.000	0.000	0.006	0.088	0.457	0.868	0.988	0.999
One sided Correlation test	0.103	0.951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: This table presents the same information provided in Table 4 with only two differences. First, Table 9 reports simulation results for DGP2 instead of DGP1. Second, the top row reports different values of the γ_Y parameter of DGP2 instead of different values for $Corr(\varepsilon^Y_{t+1}, \varepsilon^X_t)$.

Table 10: MSPE based test vs Correlation based test when T=1000, nominal size is 10%. DGP2

γ_Y	0.00	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean RMSPE benchmark	10.978	11.032	11.203	11.506	11.975	12.671	13.714	15.357	18.263	25.080
Mean RMSPE of Forecast X_t	15.527	14.804	14.173	13.617	13.121	12.676	12.274	11.908	11.572	11.264
Mean $Corr(Y_{t+1}, X_t)$	-0.002	0.098	0.198	0.298	0.397	0.497	0.597	0.696	0.796	0.896
One sided DMW test	0.000	0.000	0.000	0.000	0.000	0.100	0.984	1.000	1.000	1.000
One sided Correlation test	0.103	0.951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: This table presents the same information provided in Table 9. The only difference is that Table 10 reports simulation results when our sample size is bigger, T=1000.

The rest of the columns in Tables 9 and 10 are even more interesting because they show cases with a positive correlation between the forecast X_t and the target variable. In particular, whenever $0 < \gamma_Y \leq 0.5$ we see that forecast X_t is superior to the benchmark in terms of correlations, yet inferior in terms of MSPE. Put differently, in all this rage of values for γ_Y, the constant benchmark model is more accurate than X_t in terms of MSPE, yet Y_{t+1} is clearly predictable by X_t, as they are strongly connected with relatively high correlations. This is another example in which MSPE comparisons fail to detect predictability. The DMW does its duty and tends to correctly reject in favor of the model with lower MSPE. Similarly our correlation test does its duty also, rejecting in favor of the model with higher correlation. The problem again is that our forecast X_t is Mincer-Zarnowitz inefficient. In this scenario, traditional MSPE comparisons may fail to detect predictability as we saw earlier with DGP1 and with our basic examples in section 2.

Notice also that whenever $\gamma_Y > 0.5$ both criteria, MSPE and correlations, agree in that X_t is preferable relative to the benchmark. Moreover, both tests: DMW and our correlation test, correctly reject their corresponding null hypothesis quite frequently, reaching high power as the correlation increases.

4. Empirical Illustration

In this section we illustrate the main point of our paper with an empirical application using energy commodities and a commodity-currency: the Chilean peso. This illustration is partially inspired by the seminal papers by Chen, Rogoff and Rossi (2010, 2011, 2014) who report predictive ability from the exchange rates of commodity exporting countries to country-specific commodity indices. In the same line of argument Alquist, Kilian and Vigfusson (2013) show that commodity–currencies have an important ability to predict oil prices. Pincheira, Bentancor, Hardy and Jarsún (2022) corroborate this finding using lags of long returns of the Chilean Peso, a commodity-currency strongly and negatively correlated to copper prices[12].

Alquist and Kilian (2010) and Alquist, Kilian and Vigfuson (2013) base their forecasting models in the well known Hotelling (1931) approach, which indicates that the oil price should grow according to the interest rate. Alquist, Kilian and Vigfuson (2013) take this main idea one step further and extend Hotelling´s model to include other predictors different from the interest rate. To be more precise, they construct oil price forecasts according to the following expression:

$$P^{f}_{t+h,t} = P_t\left(1 + z_{t,h}\right) \qquad (7)$$

Where $P^{f}_{t+h\vee t}$ represents the oil price that is predicted *h* steps into the future at time *t*. P_t represents the oil price at time *t*, *h* represents the forecast horizon, whereas $z_{t,h}$ represents the percent variation of a particular exchange rate over the most recent *h* periods. While Alquist, Kilian and Vigfuson (2013) analyze the Canadian and Australian dollars, Pincheira, Bentancor, Hardy and Jarsún (2022) show that the Chilean Peso tends to outperform those currencies when predicting the oil price and a variety of oil related products. In doing so, the authors consider the following specification for their monthly data:

$$\Delta ln(P_t) = c + \beta_1[\Delta ln(ER_{t-1}) + \cdots + \Delta ln(ER_{t-3})] + \beta_2[\Delta ln(ER_{t-4}) + \cdots + \Delta ln(ER_{t-6})] + \varepsilon_t$$

where $\Delta ln(X_t) \equiv ln(X_t) - ln(X_{t-1})$ and ε_t represent an error term. P stands for "Fuel Price" and it is their target variable. Finally, ER_t represents the Chilean exchange rate at time *t*. In other words, they predict one-month-ahead oil returns with two main regressors: the three-month return of the Chilean peso from month t-3 to month t, and another three-month return of the Chilean Peso, but including months t-6 through t-4. In their Table 3 they show OLS estimates of the β_1 and β_2 coefficients. To our surprise, they are extremely similar, which suggests using directly a six-month return of the Chilean peso to predict one-month return of the oil price. Accordingly, in our empirical illustration, we build our forecasts with the following expression:

$$r^{f}_{t+1\vee t} = -r^{CLP}_{t,6} \quad (8)$$

where $r^{f}_{t+1\vee t}$ represents a one-month return forecast for a given fuel made at time t. $r^{CLP}_{t,6}$ represents a six-month log return of the Chilean peso between month t and month t-6. We multiply these returns by a minus 1 because in our database the Chilean Peso exchange rate is expressed in terms of Chilean pesos per one US dollar.

The use of long returns in commodity-currencies to predict short returns in commodity prices has been explored by a sequel of papers by Pincheira and Hardy (2019, 2021), Pincheira, Hardy, Henríquez, Tapia and Bentancor (2023) and Pincheira, Bentancor, Hardy and Jarsún (2022). This is different from Alquist, Kilian and Vigfusson (2013) who study predictability *h*-periods ahead in

[12] *Pincheira and Hardy (2019, 2021) find strong predictive ability from the Chilean Peso to some base-metal prices, including copper.*

the oil price, with *h*-period returns in commodity-currencies. The use of long returns as predictors can be seen as a way to smooth out exchange rates or to extract a stronger signal from the usually noisy short returns in forex markets. In the case of the Chilean peso, while some variation is clearly associated to international fluctuations in industrial commodity prices, some other variation is unrelated to it, and it has to do with the behavior of noisy traders or important institutional players and their own decision making process. For instance, from time to time the Central Bank of Chile either buy or sell US dollars in the local market with the specific purpose of accumulating reserves or stabilize the real exchange rate. In a similar fashion, private pension funds institutions face domestic regulations in terms of exposure to local and international risk. Consequently they frequently need to rebalance their portfolios which may importantly affect the domestic currency. These decisions might not be related to the international evolution of industrial commodity prices, and the use of long returns of the Chilean Peso may be serving as a simple filter to obtain the core signal that connects the Chilean peso with international commodity prices[13].

The source of our data is Refinitiv Datastream, from which we downloaded daily closing prices of five series: Brent oil, heating oil, propane, WTI oil and the Chilean peso (relative to the US dollar). Our data are converted to monthly frequency by sampling from the last day of the month. Our monthly database goes from January 2000 to February 2024. The starting point of our analyses is mainly determined by the date in which Chile achieved a full fledge inflation targeting regime[14]. As mentioned before, we compute one-step-ahead forecasts for fuel returns using expression (8).

We focus on two different exercises. First, we compute 283 one-step-forecasts from August 2008, to February 2024. Our first forecast uses the first six-month-return that we are able to compute with our data, going from July 2000 back to January 2000. Then we compute the correlation of our sequence of forecasts with each target variable. Notice how simple our approach is. We use the same sequence of forecasts for our four series of fuel returns. A statistical significance correlation is evaluated with our one sided R^2 test. We also compute MSPE of our forecasts with the Chilean Peso, and compare it to the MSPE of the constant zero forecast. We then evaluate the null of equality in MSPE with a standard DMW test. Table 11 shows our results although instead of raw MSPE we show RMSPE.

Table 11 illustrates the main point of our paper vividly. First, for all four fuels the RMSPE of the zero forecast is lower than the corresponding RMSPE when forecasting with the Chilean peso. Moreover, the DMW test rejects the null of equality in MSPE quite strongly in favor of the zero forecast in all four cases. In other words, the forecast based on the Chilean peso is clearly outperformed by the no change forecast in terms of MSPE. Can we conclude then that the Chilean peso has no predictive ability whatsoever for fuel returns? The last column in Table 11 has a strong negative answer to that question, as the correlation of the forecast based on the Chilean Peso with our target variables is statistically significant in all four cases. Moreover, these correlations are sizable, ranging from 0.17 for Propane, to 0.25 for WTI and Heating Oil. These statistically significant correlations indicate a strong connection between past developments in the Chilean Peso and future fuel returns. This is particularly relevant when the implicit or explicit benchmark is the no change forecast, which has zero covariance with the target variable. This is a strong case of predictability that might be overlooked by a researcher blindly focused on forecast accuracy.

Our second exercise is very compelling as well, and it is similar to the former. The main difference

---

[13] *Pincheira and Hardy (2023) show that, in many cases, exchange rate expectations have stronger ability to predict commodity returns relative to the Chilean exchange rate. The authors argue that the predictive connection between commodity-currencies and commodity returns might be better exploited with the aid of exchange rate expectations that are free from the volatility introduced by some institutional players and noisy traders as well.*

[14] *October 1999 would be the exact date.*

is that we change the benchmark. Instead of using the constant zero forecast, we use the rolling mean forecast computed with the most recent 50 observations, similar to what Goyal and Welch (2008) and Goyal, Welch and Zafirov (2023) do. As a consequence of this change in the benchmark, we are forced to use the first 50 observations in our sample to compute the first rolling mean forecast. Consequently our evaluation period is shorter now, including only 239 observations from April 2004 until February 2024. In this shorter sample, the correlation of the Chilean peso with the target variables is a little different than before. Similarly with RMSPE, they are slightly different than in our previous exercise. Despite these little differences, our main conclusions remain the same. Results are displayed in Table 12 next.

Table 11: Forecasting monthly fuel returns with the Chilean Peso. The Benchmark is the no change forecast and the sample period goes from August 2000 to February 2024

	RMSPE Zero Forecast	RMSPE CLP Forecast	RMSPE Ratio	Correlation
WTI	11.12	12.38	1.11**	0.25***
Brent	13.21	14.19	1.07**	0.22**
Propane	14.33	15.48	1.08***	0.17**
Heating Oil	10.51	11.90	1.13**	0.25***

*Notes: Table 11 presents results when forecasting monthly returns of WTI oil, Brent oil, Propane and Heating oil one- step-ahead, for the period August 2000 to February 2024. Table 11 presents RMSPE associated to the zero forecast and to another forecast constructed based on the Chilean Peso. Table 11 also presents RMSPE ratios, where a figure greater than 1 favors the zero forecast. Statistically significant differences in RMSPE are evaluated with a traditional two sided DMW test. The last column in Table 11 presents the correlation between the forecast based on the Chilean peso and our target variables. Statistically significant correlations are evaluated according to our one sided R^2 test. *p<10%, **p<5%, ***p<1%. Source: Authors' Elaboration.*

Table 12: Forecasting monthly fuel returns with the Chilean Peso. The Benchmark is the rolling mean forecast computed with the most recent 50 observations. The sample period goes from April 2004 to February 2024

	RMSPE Rolling Mean Forecast	RMSPE CLP Forecast	RMSPE Ratio	Correlation
WTI	11.54	12.44	1.08	0.26***
Brent	13.51	14.09	1.04	0.24**
Propane	12.39	13.39	1.08**	0.22***
Heating Oil	10.49	11.46	1.09	0.29***

*Notes: Table 12 presents results when forecasting monthly returns of WTI oil, Brent oil, Propane and Heating oil one- step-ahead, for the period April 2004 to February 2024. Table 12 presents RMSPE associated to the rolling mean forecast constructed with the last 50 observations of the predictand and also the RMSPE of another forecast based on the Chilean Peso. Table 12 also presents RMSPE ratios, where a figure greater than 1 favors the rolling mean forecast. Statistically significant differences in RMSPE are evaluated with a traditional two sided DMW test. The last column in Table 12 presents the correlation between the forecast based on the Chilean peso and our target variables. Statistically significant correlations are evaluated according to our one sided R^2 test. *p<10%, **p<5%, ***p<1%. Source: Authors' Elaboration.*

From Table 12 we receive the same message than from Table 11: the forecast based on the Chilean Peso is clearly outperformed in terms of forecast accuracy, with RMSPE ratios ranging between 1.04 and 1.09 in favor of the rolling mean forecast, although now the DMW test rejects the null of equality in MSPE only when predicting Propane. Despite these unfavorable results, the correlation of the forecast based on the Chilean peso and our target variables are even larger than in Table 11, ranging from 0.22 to 0.29. All four are statistically significant at least at the 5% significance level. For the curious reader, let us add that all four correlations of the rolling mean forecast with the target variable are small and negative.

One legitimate question to ask is whether our empirical results are driven by the last sample period which might suffer from important instabilities due to the COVID pandemic. In a recent paper, Iacone, Rossini and Viselli (2024) explore the behavior of global tests of predictive ability in the presence of important instabilities during short periods of time. They recommend the exclusion of those periods from traditional analyses because they might importantly alter the behavior of traditional tests evaluating average performances. We follow exactly that advice in Tables 13 and 14, where we restrict our sample to end in December 2019, before the beginning of the pandemic.

Table 13: Forecasting monthly fuel returns with the Chilean Peso excluding the COVID Pandemic. The Benchmark is the no change forecast.

	RMSPE Zero Forecast	**RMSPE CLP Forecast**	**RMSPE Ratio**	**Correlation**
WTI	9.29	10.47	1.11**	0.31***
Brent	10.37	11.35	1.07*	0.29***
Propane	14.42	15.21	1.08**	0.20**
Heating Oil	9.35	10.56	1.13**	0.31***

Notes: Table 13 is akin to Table 11 with the only difference that the sample period goes from August 2000 to December 2019, so to exclude the COVID pandemic from the analysis.

Table 14: Forecasting monthly fuel returns with the Chilean Peso excluding the COVID Pandemic. The Benchmark is the rolling mean forecast computed with the most recent 50 observations.

	RMSPE Rolling Mean Forecast	RMSPE CLP Forecast	RMSPE Ratio	Correlation
WTI	9.39	10.07	1.08	0.34***
Brent	9.97	10.41	1.04	0.35***
Propane	11.96	12.37	1.08	0.28***
Heating Oil	8.96	9.56	1.09	0.37***

Notes: Table 14 is akin to Table 12 with the only difference that the sample period goes from April 2004 to December 2019, so to exclude the COVID pandemic from the analysis.

Tables 13 and 14 corroborate the results from Tables 11 and 12 indicating that our results are not the consequence of the instabilities during the Pandemic. While there are some subtleties when comparing Table 11 with Table 13 and Table 12 with Table 14, it is striking to note that when excluding the Pandemic period, the correlations of our target variables with the forecast constructed with the Chilean peso increase, ranging from 0.20 to 0.38, being statistically significant in all cases. At the same time, this strongly correlated forecast is outperformed in terms of RMSPE by our naïve benchmarks. In fact, the null of equality in MSPE is either not rejected, or it is rejected favoring the simple benchmarks.

Consistent with the MSPE Paradox in Pincheira and Hardy (2024a), the opposite performance in correlations and MSPE is associated to the presence of inefficiencies in our forecasts. In particular, the correlation of the forecast based on the Chilean Peso and its forecast errors is always negative and statistically significant at tight significance levels. To be more precise, these correlations range between -0.37 (for Propane) to -0.53 (for WTI and Heating oil) and all of them are statistically significant at the 1% level. Tables with all the details of these inefficiencies are available upon request.

Tables 11-14 clearly illustrate that questions about forecast accuracy and predictability are not the same. When MSPE is the loss function of preference, these two questions will be different as long as at least one competing forecast violates Mincer and Zarnowitz efficiency. Several influential papers in the literature are explicit in their objective to detect if a given series is predictable or not. Yet the bulk of their analyses are forecast accuracy comparisons. In light of our simple yet strong results, the eyes should also look at measures of dependence, being correlations the natural first step in this direction.

5. Concluding remarks

Very influential articles in the forecasting literature ask a very simple question: is it a given time series Y_{t+1} predictable? Examples of papers dealing at least partially with this question are Rossi (2013), Timmermann (2008), Gargano and Timmermann (2014), Goyal and Welch (2008) and Goyal and Welch and Zafirov (2023). These are just a few examples of a vast literature exploring either directly or indirectly this question. A fairly common approach to address this issue with out-of-sample exercises, compares the MSPE of a plausible predictor for $Y_{t+1}$and the MSPE of a naïve benchmark consistent with unpredictability. Typically, the naïve benchmark is a constant forecast or some sort of historical average of the predictand. If the MSPE of the plausible predictor is lower than that of this benchmark, the series is considered predictable, otherwise it is considered unpredictable. This intuitive and standard approach, however, might not be truly capturing the essence of predictability that is perfectly summarized by Diebold and Kilian (2001) when they mention that "*The extent of a series´ predictability depends on how much information the past conveys regarding future values of this series;*" Diebold and Kilian (2001) page 657. Similarly, Clements and Hendry (1998) make use of the notion of statistical independence to define unpredictability. If we understand predictability as a connection between our target variable Y_{t+1} and information revealed in the past, then predictability and accuracy are different, albeit, related concepts. In particular, when the plausible forecast under evaluation is inefficient, it might face a paradoxical situation: On the one hand, it could have a strong and positive correlation with the target variable, much greater than the correlation of the benchmark with the same target variable. Yet, on the other hand, it could be outperformed in terms of MSPE by the same naïve benchmark.

Predictability does not require an efficient forecast. Predictability just requires dependence between the future and the set of information available at time *t*. We evaluate the performance of very simple correlation tests to detect predictability. Our Monte Carlo simulations suggest that they are well sized and relatively powerful.

We illustrate our point with an empirical application in which forecasts based on the Chilean Peso are used to predict monthly returns of oil and oil related products. Despite being strongly correlated with the respective target variable, our preferred forecast cannot outperform either the naïve rolling mean or the zero forecast in terms of MSPE.

Our empirical application clearly illustrate that questions about forecast accuracy and predictability are not equal. When MSPE is the loss function of preference, these two questions will be different as long as at least one competing forecast violates Mincer and Zarnowitz efficiency. Several influential papers in the literature are explicit in their objective to detect if a given series is predictable or not. Yet the bulk of their analyses are forecast accuracy comparisons. In light of our simple yet strong results, the eyes should also look at measures of dependence and not only at measures of forecast accuracy, being correlations the natural first step in this direction, but of course future research might also explore more complex structures of dependence, allowing for structural breaks or time varying correlations.

References

Alquist, R. and Kilian, L., 2010. What do we learn from the price of crude oil futures?. *Journal of Applied Econometrics*, 25(4), pp.539–573. https://doi.org/10.1002/jae.1159

Alquist, R., Kilian, L. and Vigfusson, R.J., 2013. Forecasting the price of oil. In: Elliott, G., ed., *Handbook of Economic Forecasting*. Vol. 2. Philadelphia, PA: Elsevier, pp.427–507.

Ang, A., Bekaert, G. and Wei, M., 2007. Do macro variables, asset markets, or surveys forecast inflation better?. *Journal of Monetary Economics*, 54(4), pp.1163–1212. https://doi.org/10.1016/j.jmoneco.2006.04.006

Chen, Y.-C., Rogoff, K. S. and Rossi, B., 2010. Can exchange rates forecast commodity prices?. *Quarterly Journal of Economics*, 125(August), pp.1145–1194. Available at: <https://www.jstor.org/stable/27867508>.

Chen, Y.-C., Rogoff, K.S. and Rossi, B., 2011. *Predicting agri-commodity prices: An asset pricing approach*. In: Munier, B., ed., World Uncertainty and the Volatility of Commodity Markets. Amsterdam: IOS Press.

Chen, Y., Rogoff, K. and Rossi, B., 2014. Can exchange rates forecast commodity prices? An update. Manuscript, February 2014. Available at: <https://scholar.harvard.edu/files/rogoff/files/crr2014a.pdf

Clark, T. E. and McCracken, M.W., 2001. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105(1), pp.85–110. https://doi.org/10.1016/S0304-4076(01)00071-9

Clark, T. E. and McCracken, M.W., 2005. Evaluating direct multistep forecasts. *Econometric Reviews*, 24(4), pp.369–404. https://doi.org/10.1080/07474930500405683

Clark, T. E. and West, K. D., 2006. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, 135(1–2), pp.155–186. https://doi.org/10.1016/j.jeconom.2005.07.014

Clark, T. E. and West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), pp.291–311. https://doi.org/10.1016/j.jeconom.2006.05.023

Clements, M. and Hendry, D., 1998. *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Diebold, F. X. and Mariano, R., 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), pp.253–263. https://doi.org/10.1080/07350015.1995.10524599

Diebold, F. X. and Kilian, L., 2001. Measuring predictability: Theory and macroeconomic applications. *Journal of Applied Econometrics*, 16(6), pp.657–669. https://doi.org/10.1002/jae.619

Gargano, A. and Timmermann, A., 2014. Forecasting commodity price indexes using macroeconomic and financial predictors. *International Journal of Forecasting*, 30(3), pp.825–843. https://doi.org/10.1016/j.ijforecast.2013.09.003

Giacomini, R. and White, H., 2006. Tests of conditional predictive ability. *Econometrica*, 74(6), pp.1545–1578. https://www.jstor.org/stable/4123083

Goyal, A. and Welch, I., 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), pp.1455–1508. https://doi.org/10.1093/rfs/hhm014

Goyal, A., Welch, I. and Zafirov, A., 2024. A comprehensive 2022 look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 37(11), pp.3490–3557. https://doi.org/10.1093/rfs/hhae044.

Hansen, B., 2022. Econometrics. Princeton: Princeton University Press.

Hotelling, H., 1931. The economics of exhaustible resources. *Journal of Political Economy*, 39(2), pp.137–175. Available at: <https://www.jstor.org/stable/1822328>.

Iacone, F., Rossini, L. and Viselli, A., 2024. Comparing predictive ability in presence of instability over a very short time. arXiv preprint, arXiv:2405.11954. Available at: <https://arxiv.org/abs/2405.11954>.

Ince, O. and 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, pp.131–151. https://doi.org/10.1016/j.intfin.2016.11.002

Joutz, F. and Stekler, H. O., 2000. An evaluation of the prediction of the Federal Reserve. *International Journal of Forecasting*, 16(1), pp.17–38. https://doi.org/10.1016/S0169-2070(99)00046-1

Mincer, J. A. and Zarnowitz, V., 1969. *The evaluation of economic forecasts*. In: Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance. NBER, pp.3–46.

Newey, W. and West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), pp.703–708. Available at: <https://www.jstor.org/stable/1913610

Newey, W. and West, K., 1994. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4), pp.631–653. https://doi.org/10.2307/2297912

Nordhaus, W., 1987. Forecasting efficiency: Concepts and applications. *The Review of Economics and Statistics*, 69(4), pp.667-674. Available at: https://www.jstor.org/stable/1935962

Patton, A. and Timmermann, A., 2012. Forecast rationality tests based on multi-horizon bounds. *Journal of Business and Economic Statistics*, 30(1), pp.1–17. https://doi.org/10.1080/07350015.2012.634337

Pesaran, M. and Timmermann, A., 1992. A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics*, 10(4), pp.461-465. Available at: <https://www.jstor.org/stable/1391822

Pesaran, M. and Timmermann, A., 2009. Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485), pp.325–337. Available at: <https://www.jstor.org/stable/40591921

Pincheira, P., 2013. Shrinkage-based tests of predictability. *Journal of Forecasting*, 32(4), pp.307–332. https://doi.org/10.1002/for.1270

Pincheira, P., 2022. A power booster factor for out-of-sample tests of predictability. *Economia*, 45(89), pp.150–183. https://doi.org/10.18800/economia.202201.006

Pincheira, P., 2010. A real-time evaluation of the Central Bank of Chile GDP growth forecasts. Money Affairs, CEMLA, 0(1), pp.37–73. Available at: <https://www.cemla.org/PDF/moneyaffairs/pub_monaff_xxiii_01.pdf>.

Pincheira, P., 2012. A joint test of superior predictive ability for Chilean inflation forecasts. Economía Chilena (The Chilean Economy), Central Bank of Chile, 15(3), pp.4–39. (In Spanish). Available at: <https://xn--economachilena-5lb.cl/index.php/economiachilena/article/view/158>.

Pincheira, P. and Hardy, N., 2019. Forecasting base metal prices with the Chilean exchange rate. *Resources Policy*, 62(February), pp.256–281. https://doi.org/10.1016/j.resourpol.2019.02.019

Pincheira, P. and Hardy, N., 2021. Forecasting aluminum prices with commodity currencies. *Resources Policy*, 73, 102066. https://doi.org/10.1016/j.resourpol.2021.102066

Pincheira, P. and Hardy, N., 2023. Forecasting base metal prices with exchange rate expectations. *Journal of Forecasting*, 42(8), pp.2341–2362. https://doi.org/10.1002/for.3018

Pincheira, P. and Hardy, N., 2024a. The mean squared prediction paradox. *Journal of Forecasting*, Early view. https://doi.org/10.1002/for.3129.

Pincheira, P. and Hardy, N., 2024b. Correlation-based tests of predictability. *Journal of Forecasting*, Early view. https://doi.org/10.1002/for.3081.

Pincheira, P., Bentancor, A. and Hardy, N., 2023. An inconvenient truth about forecast combinations. *Mathematics*, 11(18), 3806. https://doi.org/10.3390/math11183806

Pincheira-Brown, P., Bentancor, A., Hardy, N. and Jarsun, N., 2022. Forecasting fuel prices with the Chilean exchange rate: Going beyond the commodity currency hypothesis. *Energy Economics*, 106, 105802. https://doi.org/10.1016/j.eneco.2021.105802

Pincheira, P., Hardy, N., Henriquez, C., Tapia, I. and Bentancor, A., 2023. Forecasting base metal prices with an international stock index. *Czech Journal of Economics and Finance*, 73(3), pp.277–302. https://doi.org/10.32065/CJEF.2023.03.03

Rossi, B., 2013. Exchange rate predictability. *Journal of Economic Literature*, 51(4), pp.1063–1119. DOI: 10.1257/jel.51.4.1063

Timmermann, A., 2008. Elusive return predictability. *International Journal of Forecasting*, 24(1), pp.1–18. https://doi.org/10.1016/j.ijforecast.2007.07.008

West, K. D., 1996. Asymptotic inference about predictive ability. *Econometrica*, 64(5), pp.1067-1084. Available at: <https://www.jstor.org/stable/2171956

West, K. D. and Hubrich, K., 2010. Forecast comparisons in small nested model sets. *Journal of Applied Econometrics*, 25(4), pp.574–594. Available at: https://doi.org/10.1002/jae.1176

Appendix A1: Derivation of the Correlation Test

Assumptions:

Let $M_t = \begin{pmatrix} X_t \\ Y_t \\ X_t^2 \\ X_t Y_t \end{pmatrix}$, and $\widetilde{M_t} = M_t - E(M_t) = \begin{pmatrix} X_t - E(X_t) \\ Y_t - E(Y_t) \\ X_t^2 - E(X_t^2) \\ X_t Y_t - E(X_t Y_t) \end{pmatrix}$

i) The vector $\widetilde{M_t}$ is strictly stationary with mixing coefficients $\alpha(l)$ such that, for some $r > 2$, $E \parallel \widetilde{M_t} \parallel^r < \infty$ and $\sum_{l=1}^{\infty} \alpha(l)^{1-\frac{2}{r}} < \infty$

ii) A strictly positive variance for Y_t and X_t

iii) X_t is considered as "given". In other words, we do not address here the effects of parameter uncertainty.

iv) $Corr(Y, X)$ is lower than 1.

As commented in Pincheira and Hardy (2024b), assumption i) is sufficient for the CLT to hold (see Theorem 14.15 in Hansen (2022), page 470).

Correlation test of predictive ability.

Define the following sample moments vector:

$$\begin{bmatrix} m_X \\ m_Y \\ m_{XX} \\ m_{YY} \\ m_{YX} \end{bmatrix} = \frac{1}{T} \begin{bmatrix} \sum_{t=1}^{T} X_t \\ \sum_{t=1}^{T} Y_t \\ \sum_{t=1}^{T} X_t^2 \\ \sum_{t=1}^{T} Y_t^2 \\ \sum_{t=1}^{T} Y_t X_t \end{bmatrix}$$

Where $[\mu_X, \mu_Y, \mu_{XX}, \mu_{YY}, \mu_{YX}]^T$ is the population counterpart. Let $s_{YX} = m_{YX} - m_Y m_X, s_Y^2 = m_{YY} - m_Y^2, s_X^2 = m_{XX} - m_x^2$ be the sample counterparts of $Cov(Y_t, X_t) = \sigma_{YX} = \mu_{YX} - \mu_Y \mu_X, \sigma_Y^2 = \mu_{YY} - \mu_Y^2, \sigma_X^2 = \mu_{XX} - \mu_X^2$, respectively. By the CLT,

$$\sqrt{T} \left\{ \begin{bmatrix} m_X \\ m_Y \\ m_{XX} \\ m_{YY} \\ m_{YX} \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \\ \mu_{XX} \\ \mu_{YY} \\ \mu_{YX} \end{bmatrix} \right\} \xrightarrow{d} N\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \sum_{j=-\infty}^{\infty} \Omega_j \}$$

Where

$$\Omega_j = \begin{bmatrix} Cov(X_t, X_{t-j}) & Cov(Y_t, X_{t-j}) & Cov(X_t^2, X_{t-j}) & Cov(Y_t^2, X_{t-j}) & Cov(X_tY_t, X_{t-j}) \\ Cov(X_t, Y_{t-j}) & Cov(Y_t, Y_{t-j}) & Cov(X_t^2, Y_{t-j}) & Cov(Y_t^2, Y_{t-j}) & Cov(X_tY_t, Y_{t-j}) \\ Cov(X_t, X_{t-j}^2) & Cov(Y_t, X_{t-j}^2) & Cov(X_t^2, X_{t-j}^2) & Cov(Y_t^2, X_{t-j}^2) & Cov(X_tY_t, X_{t-j}^2) \\ Cov(X_t, Y_{t-j}^2) & Cov(Y_t, Y_{t-j}^2) & Cov(X_t^2, Y_{t-j}^2) & Cov(Y_t^2, Y_{t-j}^2) & Cov(X_tY_t, Y_{t-j}^2) \\ Cov(X_t, Y_{t-j}X_{t-j}) & Cov(Y_t, Y_{t-j}X_{t-j}) & Cov(X_t^2, Y_{t-j}X_{t-j}) & Cov(Y_t^2, Y_{t-j}X_{t-j}) & Cov(X_tY_t, Y_{t-j}X_{t-j}) \end{bmatrix}$$

Let us define the following function $g: R^5 \to R^3$ such that

$$g\begin{bmatrix} m_X \\ m_Y \\ m_{XX} \\ m_{YY} \\ m_{YX} \end{bmatrix} = \begin{bmatrix} s_X^2 \\ s_Y^2 \\ s_{YX} \end{bmatrix} = \begin{bmatrix} m_{XX} - m_X^2 \\ m_{YY} - m_Y^2 \\ m_{YX} - m_Y m_X \end{bmatrix}$$

Let Δg be the first derivative of g:

$$\Delta g\begin{bmatrix} m_X \\ m_Y \\ m_{XX} \\ m_{YY} \\ m_{YX} \end{bmatrix} = \begin{bmatrix} -2m_X & 0 & -m_Y \\ 0 & -2m_Y & -m_X \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Or, applied the population counterpart

$$\Delta g\begin{bmatrix} \mu_X \\ \mu_Y \\ \mu_{XX} \\ \mu_{YY} \\ \mu_{YX} \end{bmatrix} = \begin{bmatrix} -2\mu_X & 0 & -\mu_Y \\ 0 & -2\mu_Y & -\mu_X \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then, by the CLT

$$\sqrt{T}\left\{\begin{bmatrix} s_X^2 \\ s_Y^2 \\ s_{YX} \end{bmatrix} - \begin{bmatrix} \sigma_X^2 \\ \sigma_Y^2 \\ \sigma_{XY} \end{bmatrix}\right\} \underset{\rightarrow}{d} N\{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Delta g' \sum_{j=-\infty}^{\infty} \Omega_j \, \Delta g\}$$

Finally, let $h: R^3 \to R$ such that

$$h\begin{bmatrix} s_X^2 \\ s_Y^2 \\ s_{YX} \end{bmatrix} = \frac{s_{YX}}{s_Y s_X} = r$$

And let Δh be the vector of the first derivatives of h:

$$\Delta h\begin{bmatrix} \sigma_X^2 \\ \sigma_Y^2 \\ \sigma_{XY} \end{bmatrix} = \begin{bmatrix} \frac{-\rho}{2\sigma_X^2} \\ \frac{-\rho}{2\sigma_Y^2} \\ \frac{1}{\sigma_Y \sigma_X} \end{bmatrix}$$

Of course, under the null hypothesis $\rho = 0$

$$\Delta h\begin{bmatrix} \sigma_X^2 \\ \sigma_Y^2 \\ \sigma_{XY} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\sigma_Y \sigma_X} \end{bmatrix}$$

By the Delta method

$$\sqrt{T}(r-\rho)\underset{\rightarrow}{d}N\left(0,\Delta h'\Delta g'\sum_{j=-\infty}^{\infty}\Omega_j\,\Delta g\Delta h\right)$$

Then, our proposed standard normal correlation test is simply

$$Correlation-test=\frac{\sqrt{T}r}{\sqrt{\Delta h'\Delta g'\sum_{j=-\infty}^{\infty}\Omega_j\,\Delta g\Delta h}}$$

Simpler version of the test:

We start by using the same approach than in the previous lines but aimed at constructing a covariance test. Define the following sample moments vector:

$$\begin{bmatrix}m_X\\m_Y\\m_{YX}\end{bmatrix}=\frac{1}{T}\begin{bmatrix}\sum_{t=1}^{T}X_t\\\sum_{t=1}^{T}Y_t\\\sum_{t=1}^{T}Y_tX_t\end{bmatrix}$$

Where $[\mu_X,\mu_Y,\mu_{YX}]^T$ is the population counterpart. Let $s_{YX}=m_{YX}-m_Ym_X$ be the sample counterpart of $Cov(Y_t,X_t)=\sigma_{YX}=\mu_{YX}-\mu_Y\mu_X$. By the CLT

$$\sqrt{T}\left\{\begin{bmatrix}m_X\\m_Y\\m_{YX}\end{bmatrix}-\begin{bmatrix}\mu_X\\\mu_Y\\\mu_{YX}\end{bmatrix}\right\}\underset{\rightarrow}{d}N\{\begin{pmatrix}0\\0\\0\end{pmatrix},\sum_{j=-\infty}^{\infty}\Omega_j\}$$

Where $\Omega_j=\begin{bmatrix}Cov(X_t,X_{t-j}) & Cov(Y_t,X_{t-j}) & Cov(X_tY_t,X_{t-j})\\Cov(X_t,Y_{t-j}) & Cov(Y_t,Y_{t-j}) & Cov(X_tY_t,Y_{t-j})\\Cov(X_t,Y_{t-j}X_{t-j}) & Cov(Y_t,X_{t-j}Y_{t-j}) & Cov(X_tY_t,X_{t-j}Y_{t-j})\end{bmatrix}$

Let us define the following function $g\colon R^3\rightarrow R$ such that $g\begin{bmatrix}m_X\\m_Y\\m_{YX}\end{bmatrix}=s_{XY}=m_{XY}-m_Xm_Y$. Let ∇g be the first derivative of $g:\nabla g\begin{bmatrix}m_X\\m_Y\\m_{YX}\end{bmatrix}=\begin{bmatrix}-m_Y\\-m_X\\1\end{bmatrix}$. Of course, the population counterpart is just $\nabla g\begin{bmatrix}\mu_X\\\mu_Y\\\mu_{YX}\end{bmatrix}=\begin{bmatrix}-\mu_Y\\-\mu_X\\1\end{bmatrix}$. Then, by the Delta method

$$\sqrt{T}[S_{XY}-\sigma_{XY}]\underset{\rightarrow}{d}N\left(0,\nabla g'\sum_{j=-\infty}^{\infty}\Omega_j\,\nabla g\right)$$

Then, our proposed standard normal covariance test is simply

$$Covariance-test=\sqrt{T}\frac{Cov(Y,X)}{\sqrt{\nabla g'\sum_{j=-\infty}^{\infty}\Omega_j\,\nabla g}}$$

Let s_X and s_Y consistent estimates of the standard deviations of X and Y, respectively. Then our correlation test could be written as:

$$Correlation - test = \frac{\sqrt{T}\, S_{XY}/(s_X s_Y)}{(s_X s_Y)^{-1}\sqrt{\nabla g' \sum_{j=-\infty}^{\infty} \Omega_j \, \nabla g}}$$