Correlation Based Tests of Predictability

Pincheira, Pablo and Hardy, Nicolas

Universidad Adolfo Ibáñez, Universidad Diego Portales

16 February 2022

Correlation Based Tests of Predictability[1]

Pablo Pincheira Brown♣

School of Business

Universidad Adolfo Ibáñez

Nicolás Hardy§

Facultad de Economía y Empresa

Universidad Diego Portales

February 2022

Abstract

In this paper, we propose a correlation-based test for the evaluation of two competing forecasts. Under the null hypothesis of equal correlations with the target variable, we derive the asymptotic distribution of our test using the Delta method. This null hypothesis is not necessarily equivalent to the null of equal Mean Squared Prediction Errors (MSPE). Specifically, it might be the case that the forecast displaying the lowest MSPE also exhibits the lowest correlation with the target variable: this is known as "The MSPE paradox" (Pincheira and Hardy; 2021). In this sense, our approach should be seen as complementary to traditional tests of equality in MSPE. Monte Carlo simulations indicate that our test has good size and power. Finally, we illustrate the use of our test in an empirical exercise in which we compare two different inflation forecasts for a sample of OECD economies. We find more rejections of the null of equal correlations than rejections of the null of equality in MSPE.

"*The most basic form of mathematically connecting the dots between the known and unknown forms the foundations of the correlational analysis.*" Akoglu (2018).

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

Keywords: Forecasting, time-series, out-of-sample evaluation, mean squared prediction error, correlations.

[1] Corresponding Author: Pablo Pincheira-Brown

♣ Pincheira: Diagonal Las Torres 2640, Peñalolén, Santiago, Chile. Email: pablo.pincheira@uai.cl.

1. Introduction

"*The usefulness of correlation in social science research cannot be overemphasized. Establishing relationships and associations between variables, as ordinary as it may seem, does a lot to the social science researcher*." Samuel and Ethelbert (2015).

In this paper, we propose a correlation-based test for the evaluation of two competing forecasts. Under the null hypothesis of equal correlations with the target variable, we derive the asymptotic distribution of our test using the Delta method. A recent paper by Pincheira and Hardy (2021) (henceforth PH) demonstrates that the null hypothesis of equal correlations is not (necessarily) equivalent to a null hypothesis of equal MSPE[2]. In particular, they show that the forecast displaying the lowest MSPE may also exhibit the lowest correlation with the target variable: they label this result as "The MSPE Paradox." While this is an interesting observation, they do not provide formal procedures to test the null hypothesis of equal correlations. This paper fills this gap.

A crucial decision in the predictive analysis is *how* to compare our forecasts. Traditionally, the forecaster chooses some criteria, such as a loss function, to determine the *best* prediction. Nevertheless, the *best forecast* under some criteria might be the *worst* under a different criterion. For this reason, this loss function should reflect a careful analysis of the forecasting problem: Are we interested in bias? in accuracy? in comovements between time series? in a forecast-based trading rule? in something else? As pointed out by Granger and Machina (2006): "*Thus, if one forecasting method has a lower bias but higher average squared error than a second one, clients with different goals or preferences may disagree on which of the two techniques is "best" – or at least, which one is best for them*." Granger and Machina (2006), page 15. In this sense, it seems reasonable to judge forecasts under different criteria; nevertheless, as pointed out by Elliot and Timmermann (2013): "*Despite its pivotal role, it is a common practice to simply choose off-the-shelf loss functions.*" Elliot and Timmermann (2013), page 13.

Among these *off-the-shelf* loss functions, it is safe to say that MSPE is the ubiquitous criterion in the forecasting literature: "*By far the most popular loss function in empirical studies is […] mean squared error loss.*" Elliot and Timmermann (2013, page 20). Why is MSPE so popular? In our opinion, because of its simplicity and tractability: i) it is differentiable, ii) it meets some minimal properties of a "reasonable" loss function (see Granger (1999)), and iii) it is symmetric, circumventing the practical difficulty of putting weights on the relative cost of a misprediction. All in all, MSPE is an intuitive statistic that measures accuracy: the lower the MSPE, the closer the forecast will be (on average) to the target variable.

In our understanding, MSPE is popular mainly because of its simplicity, but not necessarily because it captures essential features about the utility function of the forecaster. In this regard, some assumptions in MSPE, like symmetry, are unlikely to reflect the decision-maker preferences[3]. As commented by Elliot and Timmermann (2013): "*the implicit choice of MSE loss by the majority of studies in the forecasting literature seems difficult to justify on*

[2] The authors notice that this equivalence holds exclusively under some conditions of efficiency.

[3] "[…] *an assumption of symmetry for the cost function is much less acceptable*" Granger and Newbold (1986), page 125.

economic grounds." Elliot and Timmermann (2013), page 17[4]. Along these lines, Leitch and Tanner (1991) report that professional forecasters display higher MSPE than simple time-series models; "*A natural conclusion to draw from this is that the professional forecasters objectives are poorly approximated by the MSE loss function"* Elliot and Timmermann (2013), page 37.

All in all, we think there are two important lessons from the literature. First, since the decision-maker may be interested in different features of forecasts, it seems reasonable to evaluate our predictions using different complementary approaches. Second, the popularity of those criteria relies, to some extent, on its simplicity: simple is good. In this regard, PH suggests an alternative yet straightforward approach. They suggest looking at the "association" between the prediction and the predictand rather than at "accuracy." Under this criterion, the tighter the association is, the better the prediction is.

A feature of the joint distribution of the forecast and the target variable is their covariance. In this context, PH use a simple correlation as an intuitive measure of association: "*Probably the simplest association measure between two random variables X and Y is the correlation […] a forecast more closely related to Y would be superior to another forecast not as closely related to Y. […] a forecast with higher correlation with Y should be preferable to another forecast displaying a lower correlation."* Pincheira and Hardy (2021), page 2. One problem with this approach is that the traditional theory to forecast evaluation developed by Diebold and Mariano (1995) and West (1996) does not seem to fit well when testing the null hypothesis of equal correlations with the target variable, as they comment: "*[…] an interesting avenue for future research is the elaboration of a simple asymptotically normal test to evaluate two competing forecasts according to their correlations with the target variable*." Pincheira and Hardy (2021), page 22. This paper fills this gap.

Importantly, PH argues that traditional MSPE comparisons may be somewhat misleading. When some specific conditions of efficiency are not met, the forecast displaying the lowest correlation with the target variable may also exhibit the lowest MSPE. In other words, the *less associated* forecast with the target variable may be simultaneously the *most accurate*. They show analytically, graphically, and empirically that a *useless* forecast with no relationship whatsoever with the target variable may be more accurate than a *useful* forecast displaying a positive correlation. They label this result as "*The MSPE Paradox*." While in our empirical illustration we do not find paradoxical results, we do report some remarkable differences between our correlation-based test and traditional tests comparing MSPE. In particular, we find more rejections of the null of equal correlations than rejections of the null of equality in MSPE.

A note of caution: we are not arguing that the forecast with the highest correlation should always be preferred irrespective of its MSPE. We do think, however, that approaches looking at accuracy and correlations should be seen as complementary.

[4] Granger and Machina (2006) study the utility functions implied by different loss functions. They conclude that: *"[…] utility functions associated with squared error loss are restricted to a very narrow set*."

There are four additional features of our test that are worth mentioning. First, under the null hypothesis of equal correlations, our test is asymptotically normal. Second, our Monte Carlo simulations suggest that our test is reasonably well-sized, even in multi-step-ahead exercises. Third, while the computation of our test is straightforward, it requires the estimation of a 7x7 covariance matrix. We acknowledge that some users may find this procedure cumbersome. For this reason, we also propose a "friendly-user" version of our test that works reasonably well in large samples. Finally, we emphasize that the proper environment of this paper is one in which forecasts are considered primitives: we do not address here issues arising from parameter uncertainty. In this sense, our framework is similar to Diebold and Mariano (1995)[5].

The rest of this paper is organized as follows. For completeness, in section 2, we discuss the decomposition by PH along with "The MSPE Paradox." In section 3, we present our tests and their asymptotic distributions under the null hypothesis. Section 4 outlines our experimental design and Monte Carlo simulations. To emphasize the relevance of our approach, section 5 provides an empirical illustration comparing two inflation forecasts in a sample of OECD economies. Finally, section 6 concludes.

2. The MSPE Paradox

For completeness, in this section, we illustrate what PH calls "The MSPE Paradox." The authors use this name to label the fact that when comparing two competing forecasts for the same target variable, it might be the case that the forecast displaying the lowest MSPE will also exhibit the lowest correlation with the target variable.

Let us consider $\{Y_t\}$ to be a zero-mean target variable. At time t, we have two competing forecasts $\{X_{t-1}\}$ and $\{Z_{t-1}\}$ for $\{Y_t\}$. We emphasize that both $\{X_{t-1}\}$ and $\{Z_{t-1}\}$ are forecasts constructed with information previous to time t and that they are taken as given (e.g., we are not considering issues arising from parameter uncertainty). For clarity of exposition, we drop the sub-indexes t in what follows. Let us assume that the vector (Y, X, Z) is weakly stationary and ergodic. We will also assume that both forecasts have the same non-negligible variance: Var(X)=Var(Z), that X is a zero-mean forecast, and that $E[X^2] > 0$. Many of these assumptions are very restrictive, but they are useful to illustrate the MSPE Paradox simply.

Consider now the Mean Squared Prediction Error (MSPE) of both forecasts:

$$MSPE_X = E(Y - X)^2$$

$$MSPE_Z = E(Y - Z)^2$$

[5] See West (1996) for a framework considering parameter uncertainty at the population level, Clark and McCracken (2001,2005), McCracken (2007), Clark and West (2006,2007), Pincheira and Hardy (2021) for nested models comparisons, and Giacomini and White (2006) for finite-sample predictive ability and conditional predictive ability. See West (2006) and Clark and McCracken (2013) for great reviews on forecast evaluation.

And let us also define the corresponding Mean Squared Forecasts as follows:

$$MSF_X = E[X^2]$$

$$MSF_Z = E[Z^2]$$

Let $\Delta MSPE = MSPE_X - MSPE_Z$. Appendix A1 shows that $\Delta MSPE$, *in this example,* can be decomposed as:

$$\Delta MSPE = [MSF_X - MSF_Z] - 2\sqrt{Var(Y)}\sqrt{MSF_X}\{Corr(Y,X) - Corr(Y,Z)\} \quad (1)$$

Eq. (1) illustrates a critical result: the difference in MSPE depends not only on the correlation between the forecasts with the target variable but also on the Mean Squared Forecasts (MSF). This is important since MSFs are not directly linked to the properties of the target variable.

The problem in this simple illustration relies on a "magnitude" effect that has a relevant implication on MSPE comparisons: A high MSF in a forecast could more than offset its correlation with the target variable, and therefore, underperform another less informational forecast simply because of this "magnitude" effect. In other words, *in this example*, traditional MSPE comparisons give a natural advantage to "small forecasts."

Let us now drop our simplifying assumptions and consider a more general form to explore this MSPE Paradox. Additionally, let $EX = \mu_x$, $EZ = \mu_z$ and $EY = \mu_y$. In Appendix A2 we show that $\Delta MSPE$ can be decomposed as:

$$\Delta MSPE = MSF_X - MSF_Z - 2\sqrt{V(Y)}\left\{Corr(Y,X)\sqrt{MSF_x - \mu_x^2} - Corr(Y,Z)\sqrt{MSF_z - \mu_z^2}\right\} - 2\{\mu_y(\mu_x - \mu_z)\} \quad (2)$$

Differing from our previous example, where the Paradox emerges entirely by the "magnitude" effect, in the more general case of eq.(2), the Paradox may also arise as a consequence of a complex interaction of all the terms involved in that expression. Interestingly, in contrast to eq.(1), the elimination of the "magnitude effect" ($MSF_X = MSF_Z$) is not a sufficient condition to solve the Paradox. Even if we circumvent the "magnitude" problem, we could still find that the forecast with the lowest MSPE has also the lowest correlation. As an illustration, suppose $Corr(Y,X) > Corr(Y,Z)$, $MSF_X = MSF_Z$ and $\mu_z > \mu_x$, then

$$\Delta MSPE = -2\sqrt{V(Y)}\left\{Corr(Y,X)\sqrt{MSF_x - \mu_x^2} - Corr(Y,Z)\sqrt{MSF_x - \mu_z^2}\right\} - 2\{\mu_y(\mu_x - \mu_z)\}.$$

Notice that we could still find the Paradox whenever $|\mu_y(\mu_x - \mu_z)| > \sqrt{V(Y)}\left\{Corr(Y,X)\sqrt{MSF_x - \mu_x^2} - Corr(Y,Z)\sqrt{MSF_x - \mu_z^2}\right\}$.

Propositions 1 and 2 in PH offer a characterization of "Paradox zones": they find conditions under which we may observe the Paradox. While these conditions depend on a complex interaction of different terms, the authors offer simple simulations to show that these "Paradox zones" are, in general, non-empty sets. Moreover, they show that we may have an extreme case in which an uncorrelated forecast with the target variable could be superior in terms of MSPE to an alternative forecast displaying a positive correlation with the same target variable.

3. Our correlation-based test for two competing forecasts

3.1 The Correlation test

Our test evaluates the correlation of two competing forecasts (Z and X, both with *positive variance*) with a target variable Y. For clarity of exposition, we are not using sub-indices to indicate the forecast horizon; nevertheless, we emphasize that our test is helpful for both one-step and multi-step ahead forecasts.

Let $M = M_t = \begin{pmatrix} Z_t \\ X_t \\ Y_t \\ Z_t^2 \\ X_t^2 \\ Z_t Y_t \\ X_t Y_t \end{pmatrix} = \begin{pmatrix} Z \\ X \\ Y \\ Z^2 \\ X^2 \\ ZY \\ XY \end{pmatrix}$, and $\widetilde{M} = M - E(M) = \begin{pmatrix} Z - E(Z) \\ X - E(X) \\ Y - E(Y) \\ Z^2 - E(Z^2) \\ X^2 - E(X^2) \\ ZY - E(ZY) \\ XY - E(XY) \end{pmatrix}$

Our main assumptions are the following:

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

ii) A strictly positive variance for Y, X, and Z.

iii) X and Z are considered as primitives (i.e., we do not address here the effects of parameter uncertainty).

iv) Corr(Y,X) and Corr (Y,Z) are both strictly lower than 1.

Notice that i) is sufficient conditions for the CLT to hold, see Theorem 5.20 in White (2000). See also Theorem 14.19 in Hansen (2010).

We emphasize that these assumptions are similar to those in Pincheira and Hardy (2021), with a crucial difference: both forecasts must display a strictly positive variance. This condition rules out the zero-forecast.

Let ρ_{i} be the correlation between forecast "i" and the target variable Y, with i={Z, X}. We are interested in the null hypothesis $H_0: \rho_{\mathrm{z}} = \rho_{\mathrm{x}}$. Under this null, both forecasts have the same correlation with the target variable Y. We consider the following t-statistic

$$Correlation - t = \sqrt{T s_y^2}\,\left(\frac{r_{\mathrm{z}} - r_{\mathrm{x}}}{\sqrt{\hat{V}}}\right)$$

Where $\hat{V} = \nabla\hat{h}'\nabla\hat{g}'[\sum_{j=-\infty}^{\infty} \widehat{\Omega}_{\mathrm{j}}]\,\nabla\hat{g}\nabla\hat{h}$, r_{z} and $\mathrm{r_x}$ stands for the sample correlations of Z and X with Y, respectively. Notice that s_y^2 is the sample variance of the target variable, T is the number of forecasts, $\nabla\hat{h}' = [-\frac{1}{2}(s_{yz}/\,s_z^3);\ \frac{1}{2}(s_{yx}/s_x^3);\ 1/s_z;\ 1/s_x]$, where s_z, s_x, s_{yx} and s_{yz} denote the sample standard deviations and covariances, respectively,

$$\nabla\hat{g} = \begin{pmatrix} -2m_z & 0 & -m_y & 0 \\ 0 & -2m_x & 0 & -m_y \\ 0 & 0 & -m_z & -m_x \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

Where m_z, m_x and m_y stand for the sample mean of Z, X, and Y respectively and $\Omega_{\mathrm{j}} =$

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

We show that, under the null hypothesis, our correlation-based statistic has a standard normal distribution. The long-run variance $\sum_{j=-\infty}^{\infty} \hat{\Omega}_{\mathrm{j}}$ can be estimated with a HAC estimator (e.g., Newey-West (1987, 1994), and Andrews (1991)). See Appendix A.4 for details on the formal derivation of this test.

3.2 The "friendly-user" correlation test

Even though our test is straightforward, it requires the estimation of a 7x7 matrix $\sum_{j=-\infty}^{\infty} \Omega_{\mathrm{j}}$. Our test could be further simplified if we are willing to assume $EY = EZ = EX = 0$[6]. In this case, we present a "friendly user" version of our test. Using the Delta method, we show that our correlation-based test is

$$Friendly\ Correlation - t = \sqrt{Ts_y^2}\,\left(\frac{r_{\mathrm{z}} - r_{\mathrm{x}}}{\sqrt{\widehat{W}}}\right)$$

Where the variance $\widehat{W} = \nabla\hat{h}'[\sum_{j=-\infty}^{\infty} \hat{\Gamma}_{\mathrm{j}}]\, \nabla\hat{h}$ is determined by $\nabla h' = \left(-\frac{s_{yz}}{2s_z^3}, \frac{s_{yx}}{2s_x^3}, \frac{1}{s_z}, -\frac{1}{s_x}\right)$, and

$$\Gamma_{\mathrm{j}} = \begin{pmatrix} Cov(Z_t^2, Z_{t-j}^2) & Cov(X_t^2, Z_{t-j}^2) & Cov(Z_tY_t, Z_{t-j}^2) & Cov(X_tY_t, Z_{t-j}^2) \\ Cov(Z_t^2, X_{t-j}^2) & Cov(X_t^2, X_{t-j}^2) & Cov(Z_tY_t, X_{t-j}^2) & Cov(X_tY_t, X_{t-j}^2) \\ Cov(Z_t^2, Z_{t-j}Y_{t-j}) & Cov(X_t^2, Z_{t-j}Y_{t-j}) & Cov(Z_tY_t, Z_{t-j}Y_{t-j}) & Cov(X_tY_t, Z_{t-j}Y_{t-j}) \\ Cov(Z_t^2, X_{t-j}Y_{t-j}) & Cov(Z_t^2, X_{t-j}Y_{t-j}) & Cov(Z_tY_t, X_{t-j}Y_{t-j}) & Cov(X_tY_t, X_{t-j}Y_{t-j}) \end{pmatrix}$$

Where the long-run variance $\sum_{j=-\infty}^{\infty} \Gamma_{\mathrm{j}}$ can be estimated with a HAC estimator. Under the null hypothesis of equal correlations, this statistic has a standard normal distribution. See Appendix A.4 for a formal derivation.

3.3 The case of Auto-Efficiency and unbiasedness

[6] These assumptions could be reasonable if we subtract the sample mean of each forecast and the target variable.

One implication of the MSPE Paradox is that a null hypothesis of equal correlations may differ from a null hypothesis of equal MSPE. Nevertheless, Proposition 3 in PH shows that the Paradox arises as a consequence of inefficiencies. To illustrate, let $u_x = Y - X$ and $u_z = Y - Z$ be the forecast errors of X and Z, respectively. Let us recall that X and Z are efficient *à la* Mincer and Zarnowitz (1969) as long as

$$Cov(u_x, X) = Cov(u_z, Z) = 0\ (Auto - Efficiency)$$

And

$$Eu_x = Eu_z = 0\ (unbiasedness)$$

Then, Proposition 3 in PH establishes: "*If X and Z are both efficient à la Mincer and Zarnowitz, then the Paradox is impossible*."

Unbiasedness and Auto-Efficiency are considered to be features of a "rational" forecaster. Nevertheless, these conditions do not seem to hold in practice very often. For instance, for the case of exchange rates expectations, Ince and Molodtsova (2017) report a strong rejection of them for nine developed economies (plus the euro area), and most of the 23 emerging economies analyzed. Furthermore, rejections of some properties related to optimal forecasts are also reported by Ang, Bekaert and Wei (2007); Joutz and Stekler (2000); Bentancor and Pincheira (2010); Nordhaus (1987); Patton and Timmermann (2012); Pincheira and Fernández (2011); Pincheira (2012, 2010) and Pincheira and Álvarez (2009) just to mention a few. In this sense, MSPE and correlations should be seen as complementary approaches rather than interchangeable analyses.

4 Simulations

For our simulations, we define both a data generating process (DGP) and a forecast generating process (FGP). In what follows, we assume stationarity in every process.

Let Y_t be our target variable. Our VAR(1) DGP is mainly inspired by Busetti and Marcucci (2013). Consider the target variable Y_t generated by an AR(1) component, and an unobservable variable X_{t-1}.

$$Y_t = \mu_y + \phi_y Y_{t-1} + cX_{t-1} + \varepsilon_t \quad (3)$$

$$X_t = \mu_x + \phi_x X_{t-1} + u_t \quad (4)$$

By assumption X_t follows a stationary AR(1) process. We also assume that the pair $(\varepsilon_t, u_t)'$ is a white noise vector such that

$$\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} \sim N(0, \begin{pmatrix} \sigma_\varepsilon & \rho_{\varepsilon,u}\sigma_u \\ \rho_{\varepsilon,u}\sigma_u & \sigma_u^2 \end{pmatrix})$$

Even though X_t is unobservable to the forecaster at time t, we will assume that two reasonable observable proxies for X_t are available: $Z_{1,t}$ and $Z_{2,t}$. This setup is similar to what a macroeconomist will face when dealing with different measures of core inflation or output gap, variables that can be approximated in different alternative ways.

In our simulations, we generate our proxies $Z_{1,t}$ and $Z_{2,t}$ as follows

$$Z_{1,t} = \alpha_1 + \beta_1 X_t + \omega_t \quad (5)$$

$$Z_{2,t} = \alpha_2 + \beta_2 X_t + v_t \quad (6)$$

where the pair $(\omega_t, v_t)'$ is also a Gaussian white noise vector that is totally independent from $(\varepsilon_t, u_t)'$.

We consider the following two competing h-steps-ahead forecasts for Y_{t+h-1}:

$$Y_{1,t-1}^{f}(h) = \mu_y + \phi_y Y_{t-1} + cZ_{1,t-1} \quad (FGP_1)$$

$$Y_{2,t-1}^{f}(h) = \mu_y + \phi_y Y_{t-1} + cZ_{2,t-1} \quad (FGP_2)$$

To avoid excessive notation, here we have assumed that both forecasts are built using the same population parameters μ_y, ϕ_y and c that define our main DGP. In the Appendix section we show simulations relaxing this simplifying assumption. Qualitatively speaking, results are similar.

To evaluate the empirical size of our test, we impose the null hypothesis of equal correlation with the target variable $H_0: Corr\left(Y_{1,t-1}^{f}(h), Y_{t+h-1}\right) = Corr\left(Y_{2,t-1}^{f}(h), Y_{t+h-1}\right)$ by choosing specific values of the following parameters: $\phi_y, c, \phi_x, \rho_{\varepsilon,u}, \sigma_\varepsilon, \sigma_u, \beta_1, \beta_2, V(\omega_t)$ and $V(v_t)$. To that end, we make use of the following straightforward results that stem from the definition of our DGP, FGP_1 and FGP_2:

From the fact that X_t is an AR(1) process, it follows that

$$V(X_t) = \frac{\sigma_u^2}{1-\phi_x^2}$$

From equations (3) and (4):

$$Cov(Y_t, X_t) = \frac{\phi_x c V(X_t) + Cov(u_t, \varepsilon_t)}{1-\phi_x \phi_y}$$

$$V(Y_t) = \frac{c^2 V(X_t) + 2\phi_y c Cov(Y_t, X_t) + V(\varepsilon_t)}{1-\phi_y^2}$$

Using equation (3) together with equations (5) and (6):

$$Cov(Z_{1t}, Y_t) = \beta_1 Cov(X_t, Y_t)$$

$$Cov(Z_{2t}, Y_t) = \beta_2 Cov(X_t, Y_t)$$

Similarly, using equation (4) together with equations (5) and (6):

$$Cov(Z_{1t}, X_t) = \beta_1 V(X_t)$$

$$Cov(Z_{2t}, X_t) = \beta_2 V(X_t)$$

Notice that the variances of FGP1 and FGP2 are given by

$$V\left(Y_{1,t-1}^f(h)\right) = \phi_y^2 V(Y_t) + c^2\left(\beta_1^2 V(X_t) + V(\omega_t)\right) + 2\phi_y c Cov\left(Z_{1,t}, Y_t\right)$$

$$V\left(Y_{2,t-1}^f(h)\right) = \phi_y^2 V(Y_t) + c^2\left(\beta_2^2 V(X_t) + V(v_t)\right) + 2\phi_y c Cov\left(Z_{2,t}, Y_t\right)$$

Hence the covariances of our one-step-ahead forecasts with the target variable are given by:

$$Cov\left(Y_{1,t-1}^f(1), Y_t\right) = \phi_y^2 V(Y_t) + c\phi_y Cov\left(Y_t, Z_{1,t}\right) + c\phi_y Cov(X_t, Y_t) + c^2 Cov\left(X_t, Z_{1,t}\right)$$

$$Cov\left(Y_{2,t-1}^f(1), Y_t\right) = \phi_y^2 V(Y_t) + c\phi_y Cov\left(Y_t, Z_{2,t}\right) + c\phi_y Cov(X_t, Y_t) + c^2 Cov\left(X_t, Z_{2,t}\right)$$

And the corresponding correlations are given by

$$Corr\left(Y_{1,t-1}^f(1), Y_t\right) = \frac{Cov\left(Y_{1,t-1}^f(1), Y_t\right)}{\sqrt{V\left(Y_{1,t-1}^f(1)\right)V(Y_t)}} =$$

$$\frac{\frac{\sigma_u^2 c^2}{1-\phi_x^2}\left(\beta_1 + \frac{\phi_y^2}{1-\phi_y^2}\right) + \sigma_u c\phi_y\left(\frac{\frac{\sigma_u}{1-\phi_x^2}\phi_x c + \rho_{\varepsilon,u}\sigma_\varepsilon}{1-\phi_x\phi_y}\right)\left(\frac{2\phi_y^2}{1-\phi_y^2} + \beta_1 + 1\right) + \frac{\phi_y^2\sigma_\varepsilon^2}{1-\phi_y^2}}{\sqrt{\left(\frac{c^2\sigma_u^2\left(1+\frac{2\phi_y\phi_x}{1-\phi_y\phi_x}\right)}{(1-\phi_x^2)(1-\phi_y^2)} + \frac{2\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u\phi_y c}{(1-\phi_y^2)(1-\phi_y\phi_x)} + \frac{\sigma_\varepsilon^2}{1-\phi_y^2}\right)\left[\frac{\sigma_u^2 c^2\left(\frac{\phi_y^2}{1-\phi_y^2} + \beta_1^2\left(1+\frac{2\phi_y\phi_x}{1-\phi_y\phi_x}\right) + \frac{2\phi_y^3\phi_x}{(1-\phi_y^2)(1-\phi_y\phi_x)}\right)}{1-\phi_x^2} + \frac{2\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u\phi_y c\left(\frac{\phi_y^2}{1-\phi_y^2}+\beta_1\right)}{1-\phi_y\phi_x} + \frac{\phi_y^2\sigma_\varepsilon^2}{1-\phi_y^2} + c^2V(\omega_t)\right]}} \quad (7)$$

$$Corr\left(Y_{2,t-1}^f(1), Y_t\right) = \frac{Cov\left(Y_{2,t-1}^f(1), Y_t\right)}{\sqrt{V\left(Y_{2,t-1}^f(1)\right)V(Y_t)}} =$$

$$\frac{\frac{\sigma_u^2 c^2}{1-\phi_x^2}\left(\beta_2 + \frac{\phi_y^2}{1-\phi_y^2}\right) + \sigma_u c\phi_y\left(\frac{\frac{\sigma_u}{1-\phi_x^2}\phi_x c + \rho_{\varepsilon,u}\sigma_\varepsilon}{1-\phi_x\phi_y}\right)\left(\frac{2\phi_y^2}{1-\phi_y^2} + \beta_2 + 1\right) + \frac{\phi_y^2\sigma_\varepsilon^2}{1-\phi_y^2}}{\sqrt{\left(\frac{c^2\sigma_u^2\left(1+\frac{2\phi_y\phi_x}{1-\phi_y\phi_x}\right)}{(1-\phi_x^2)(1-\phi_y^2)} + \frac{2\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u\phi_y c}{(1-\phi_y^2)(1-\phi_y\phi_x)} + \frac{\sigma_\varepsilon^2}{1-\phi_y^2}\right)\left[\frac{\sigma_u^2 c^2\left(\frac{\phi_y^2}{1-\phi_y^2} + \beta_2^2\left(1+\frac{2\phi_y\phi_x}{1-\phi_y\phi_x}\right) + \frac{2\phi_y^3\phi_x}{(1-\phi_y^2)(1-\phi_y\phi_x)}\right)}{1-\phi_x^2} + \frac{2\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u\phi_y c\left(\frac{\phi_y^2}{1-\phi_y^2}+\beta_2\right)}{1-\phi_y\phi_x} + \frac{\phi_y^2\sigma_\varepsilon^2}{1-\phi_y^2} + c^2V(v_t)\right]}} \quad (8)$$

For h-steps-ahead forecasts, we can iterate forward equation (3) to the following equations

$$Y_t = \mu_y + \phi_y Y_{t-1} + cX_{t-1} + \varepsilon_t$$

$$Y_{t+1} = \mu_y\left(1+\phi_y\right) + c\mu_x + \phi_y^2 Y_{t-1} + c(\phi_y + \phi_x)X_{t-1} + \varepsilon_{t+1} + \phi_y\varepsilon_t + cu_t$$

$$Y_{t+2} = \mu_y\left(1+\phi_y+\phi_y^2\right) + c\mu_x\left(1+\phi_y+\phi_x\right) + \phi_y^3 Y_{t-1} + c\left[\phi_y^2 + \left(\phi_y+\phi_x\right)\phi_x\right]X_{t-1} + \varepsilon_{t+2} + \phi_y\varepsilon_{t+1} + \phi_y^2\varepsilon_t + cu_{t+1} + c\left(\phi_y+\phi_x\right)u_t$$

$$Y_{t+h} = \mu_y \sum_{j=0}^{h} \phi_y^j + \sum_{j=0}^{h} \phi_y^{h-j} \varepsilon_{t+j} + \phi_y^{h+1} Y_{t-1} + \left(\sum_{j=0}^{h} \phi_y^j \phi_x^{h-j} \right) c X_{t-1} + c\mu_x \sum_{j=0}^{h-1} \sum_{i=0}^{j} \phi_y^i \phi_x^{j-i} + c \sum_{j=0}^{h-1} u_{t+h-1-j} \sum_{i=0}^{j} \phi_y^i \phi_x^{j-i}, \forall h \geq 1$$

Hence the covariances of our h-steps-ahead forecasts with the target variable are given by:

$$Cov\left(Y_{t+h-1}; Y_{1,t-1}^f(h)\right) = \phi_y^{h+1} V(Y_t) + \phi_y c \left[\left(\sum_{j=0}^{h-1} \phi_y^j \phi_x^{h-1-j} \right) Cov(Y_t; X_t) + \phi_y^{h-1} Cov\left(Y_t; Z_{1,t}\right) \right] + c^2 \left(\sum_{j=0}^{h-1} \phi_y^j \phi_x^{h-1-j} \right) Cov(X_t; Z_{1,t})$$

$$Cov\left(Y_{t+h-1}; Y_{2,t-1}^f(h)\right) = \phi_y^{h+1} V(Y_t) + \phi_y c \left[\left(\sum_{j=0}^{h-1} \phi_y^j \phi_x^{h-1-j} \right) Cov(Y_t; X_t) + \phi_y^{h-1} Cov\left(Y_t; Z_{2,t}\right) \right] + c^2 \left(\sum_{j=0}^{h-1} \phi_y^j \phi_x^{h-1-j} \right) Cov(X_t; Z_{2,t})$$

And the corresponding correlations are given by

$$Corr\left(Y_{1,t-1}^f(h), Y_{t+h-1}\right) = \frac{Cov\left(Y_{1,t-1}^f(h), Y_{t+h-1}\right)}{\sqrt{V\left(Y_{1,t-1}^f(h)\right) V(Y_t)}} =$$

$$\frac{\phi_y^{h+1}\left(\frac{c^2\sigma_u^2}{(1-\phi_x^2)(1-\phi_y^2)}\left(1+\frac{2\phi_y\phi_x}{1-\phi_y\phi_x}\right)+\frac{2\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u\phi_y c}{(1-\phi_y^2)(1-\phi_y\phi_x)}+\frac{\sigma_\varepsilon^2}{1-\phi_y^2}\right)+\phi_y c\left[\frac{\phi_x c\frac{\sigma_u^2}{1-\phi_x^2}+\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y}\left(\left(\sum_{j=0}^{h-1}\phi_y^j\phi_x^{h-1-j}\right)+\phi_y^{h-1}\beta_1\right)\right]+c^2\left(\sum_{j=0}^{h-1}\phi_y^j\phi_x^{h-1-j}\right)\beta_1\frac{\sigma_u^2}{1-\phi_x^2}}{\sqrt{\left(\frac{c^2\sigma_u^2\left(1+\frac{2\phi_y\phi_x}{1-\phi_y\phi_x}\right)}{(1-\phi_x^2)(1-\phi_y^2)}+\frac{2\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u\phi_y c}{(1-\phi_y^2)(1-\phi_y\phi_x)}+\frac{\sigma_\varepsilon^2}{1-\phi_y^2}\right)\left[\frac{\sigma_u^2 c^2\left(\frac{\phi_y^2}{1-\phi_y^2}+\beta_1^2\left(1+\frac{2\phi_y\phi_x}{1-\phi_y\phi_x}\right)+\frac{2\phi_y^3\phi_x}{(1-\phi_y^2)(1-\phi_y\phi_x)}\right)}{1-\phi_x^2}+\frac{2\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u\phi_y c\left(\frac{\phi_y^2}{1-\phi_y^2}+\beta_1\right)}{1-\phi_y\phi_x}+\frac{\phi_y^2\sigma_\varepsilon^2}{1-\phi_y^2}+c^2V(\omega_t)\right]}} \quad (9)$$

$$Corr\left(Y_{2,t-1}^f(h), Y_{t+h-1}\right) = \frac{Cov\left(Y_{2,t-1}^f(h), Y_{t+h-1}\right)}{\sqrt{V\left(Y_{2,t-1}^f(h)\right) V(Y_t)}} =$$

$$\frac{\phi_y^{h+1}\left(\frac{c^2\sigma_u^2}{(1-\phi_x^2)(1-\phi_y^2)}\left(1+\frac{2\phi_y\phi_x}{1-\phi_y\phi_x}\right)+\frac{2\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u\phi_y c}{(1-\phi_y^2)(1-\phi_y\phi_x)}+\frac{\sigma_\varepsilon^2}{1-\phi_y^2}\right)+\phi_y c\left[\frac{\phi_x c\frac{\sigma_u^2}{1-\phi_x^2}+\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y}\left(\left(\sum_{j=0}^{h-1}\phi_y^j\phi_x^{h-1-j}\right)+\phi_y^{h-1}\beta_2\right)\right]+c^2\left(\sum_{j=0}^{h-1}\phi_y^j\phi_x^{h-1-j}\right)\beta_2\frac{\sigma_u^2}{1-\phi_x^2}}{\sqrt{\left(\frac{c^2\sigma_u^2\left(1+\frac{2\phi_y\phi_x}{1-\phi_y\phi_x}\right)}{(1-\phi_x^2)(1-\phi_y^2)}+\frac{2\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u\phi_y c}{(1-\phi_y^2)(1-\phi_y\phi_x)}+\frac{\sigma_\varepsilon^2}{1-\phi_y^2}\right)\left[\frac{\sigma_u^2 c^2\left(\frac{\phi_y^2}{1-\phi_y^2}+\beta_2^2\left(1+\frac{2\phi_y\phi_x}{1-\phi_y\phi_x}\right)+\frac{2\phi_y^3\phi_x}{(1-\phi_y^2)(1-\phi_y\phi_x)}\right)}{1-\phi_x^2}+\frac{2\rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u\phi_y c\left(\frac{\phi_y^2}{1-\phi_y^2}+\beta_2\right)}{1-\phi_y\phi_x}+\frac{\phi_y^2\sigma_\varepsilon^2}{1-\phi_y^2}+c^2V(v_t)\right]}} \quad (10)$$

4.1 Size analysis: one-step-ahead forecasts and Gaussian innovations

Here we impose the null hypothesis at different levels of correlations. The idea is to choose specific values for the parameters $\phi_y, c, \phi_x, \rho_{\varepsilon,u}, \sigma_\varepsilon, \sigma_u, \beta_1, \beta_2, V(\omega_t)$ and $V(v_t)$, to equalize expressions (7) and (8). We consider three scenarios: 1) A low-correlation scenario ($Corr\left(Y_{1,t-1}^f(1), Y_t\right) = Corr\left(Y_{2,t-1}^f(1), Y_t\right) = 0.118$), 2) A mid-correlation scenario ($Corr\left(Y^f_{1,t-1}(1), Y_t\right) = Corr\left(Y^f_{2,t-1}(1), Y_t\right) = 0.687$) and 3) A high-correlation scenario ($Corr\left(Y^f_{1,t-1}(1), Y_t\right) = Corr\left(Y^f_{2,t-1}(1), Y_t\right) = 0.812$).

For the low-correlation scenario, we set the following parameters: $\phi_y = 0.15, c = 0.2, \phi_x = 0.1, \sigma^2_\varepsilon = 1, \sigma^2_u = 1.5, Corr(\varepsilon_t, u_t) = 0, \beta_1 = 0.1, \beta_2 = 0.248, \sigma^2_\omega = 1, \sigma^2_v = 2, \mu_y = 0.2, \mu_x = 0.3, \alpha_1 = 0.1$ and $\alpha_2 = 0.2$. With these parameters both correlations $Corr\left(Y^f_{1,t-1}(1), Y_t\right)$ and $Corr\left(Y^f_{2,t-1}(1), Y_t\right)$ take the same value of 0.118 and therefore the null hypothesis is satisfied.

In the mid-correlation scenario, we set: $\phi_y = 0.5, c = 0.7, \phi_x = 0.4, \sigma^2_u = 1.5, \sigma^2_\varepsilon = 1, Corr(\varepsilon_t, u_t) = 0.5, \beta_1 = 0.6, \beta_2 = 0.77, \sigma^2_\omega = 2,$ $\sigma^2_v = 2.5, \mu_y = 0.3, \mu_x = 0.4, \alpha_1 = 0.1$ and $\alpha_2 = 0.2$. In this case, both correlations $Corr\left(Y^f_{1,t-1}(1), Y_t\right)$ and $Corr\left(Y^f_{2,t-1}(1), Y_t\right)$ take the same value of 0.687 and again the null hypothesis is satisfied.

Finally, in the high-correlation scenario, we set: $\phi_y = 0.65, c = 0.65, \phi_x = 0.65, \sigma^2_\varepsilon = 1, \sigma^2_u = 1, Corr(\varepsilon_t, u_t) = 0, \beta_1 = 0.445, \beta_2 = 0.617, \sigma^2_\omega = 1,$ $\sigma^2_v = 1.3, \mu_y = 0.1, \mu_x = 0.1, \alpha_1 = 0.1$ and $\alpha_2 = 0.1$. With these parameters, $Corr\left(Y^f_{1,t-1}(1), Y_t\right) = Corr\left(Y^f_{2,t-1}(1), Y_t\right) = 0.812$ and the null hypothesis is imposed again.

Table 1 reports our results for these three scenarios using nominal sizes of 10% and 5%. We show results for our correlation test and the simplified "friendly user" version of it[7]. We consider 10,000 Monte Carlo simulations of 4 different samples sizes T = 50, 100, 500 and 2000. We estimate the long-run variances using a Barlett kernel with an automatic selection of the lag truncation parameter following Newey and West (1987,1994).

Table 1: Size analysis imposing the null hypothesis $H_0: Corr\left(Y^f_{1,t-1}(1), Y_t\right) = Corr\left(Y^f_{2,t-1}(1), Y_t\right)$ for three different scenarios of correlations.

(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Correlation scenario	Low-Correlation		Mid-Correlation		High-Correlation	
Nominal Size	*10%*	*5%*	*10%*	*5%*	*10%*	*5%*
			Sample Size=2000			
Correlation test	9.91	5.11	10.05	4.99	10.28	5.27
Friendly-user test	10.36	5.56	10.32	5.25	10.30	5.35
			Sample Size=500			
Correlation test	10.06	5.04	10.34	5.19	10.55	5.30
Friendly-user test	10.95	5.80	10.94	5.70	10.70	5.42
			Sample Size=100			
Correlation test	10.32	4.84	10.78	5.23	11.49	6.07
Friendly-user test	12.83	7.09	12.71	7.05	12.25	6.62
			Sample Size=50			
Correlation test	10.37	4.52	10.53	4.97	12.45	6.28
Friendly-user test	14.90	9.01	14.77	8.49	14.27	7.84

[7] As commented above, to implement the "friendly-user" test, we substract the sample means in each iteration.

Notes: Each entry represents the percentage of rejections of the null hypothesis of equal correlations. Columns (2), (4) and (6) consider a nominal size of 10%, while columns (3), (5) and (7) consider a nominal size of 5%. As both versions of our test are asymptotically normal, we reject the null hypothesis using standard normal critical values. We consider 10,000 Monte Carlo simulations for each exercise. The "friendly user" version of our test requires subtracting the mean of both forecasts and the target variable. Long-run variances are estimated using Newey and West (1987, 1994). Source: Author's elaboration.

Table 1 exhibits some interesting features. First, both versions of our test are reasonably well-sized for large samples (T≥ 500): notably, the size of our test does not deteriorate with a higher correlation under the null hypothesis. In large samples, the empirical size of our "Correlation test" ranges from 9.91% (4.99%) to 10.55% (5.30%) for a nominal size of 10% (5%), with a large-sample average size of 10.22% (5.15%). Similarly, for our "friendly-user test" the empirical size ranges from 10.30% (5.25%) to 10.95 (5.80%), with an average size across these six exercises of 10.60% (5.51%).

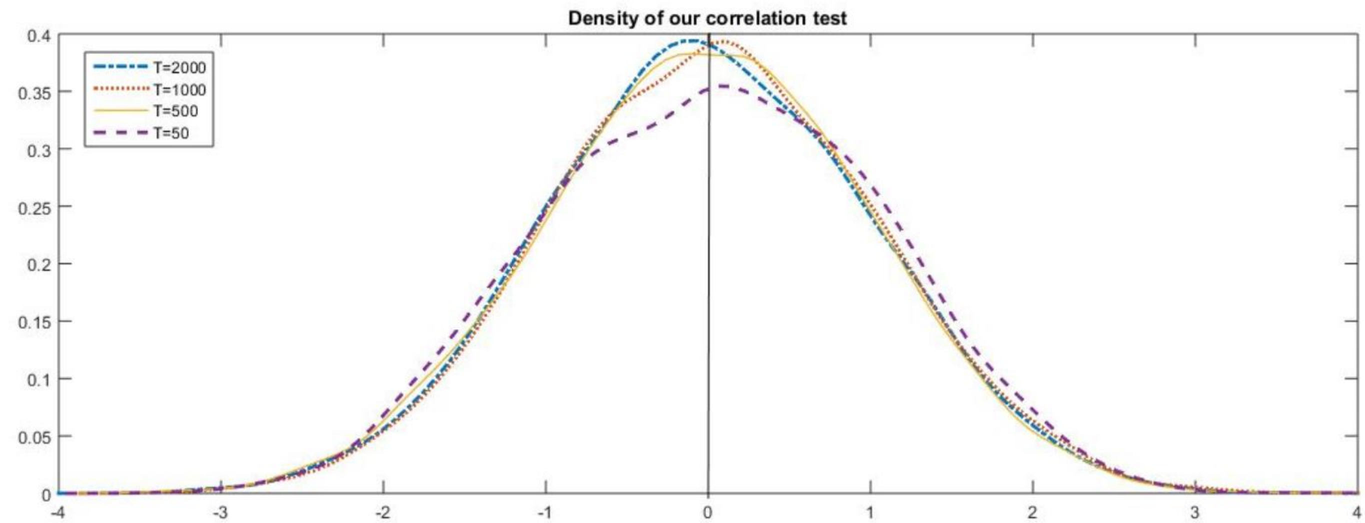
Second, our test becomes somewhat oversized in small samples (T≤ 100). The rejections rates for our "Correlation test" range from 10.32% (4.52%) to 12.45% (6.28%). While these empirical sizes may be reasonable for a small sample, the "friendly-user test" results are less encouraging. In particular, rejections rates range from 12.25% (6.62%) to 14.90% (9.01%).

Third, we do not see a clear pattern that relates size distortions and the level of correlations under the null hypothesis. While the best (worst) results for our "Correlation test" are found in the low (high)-correlation scenario, these differences tend to be negligible: Even in the high-correlations scenario, the average size of our "Correlation test" is 11.19% (5.73%), with excellent results in large samples, as expected. Of course, the persistence of the DGP may explain some of these minor distortions: the high-correlation scenario coincides with a higher persistence of our DGP. Interestingly, in the "friendly-user test," we observe cases in which the low-correlation scenario exhibits the worsts results (although with tiny differences).

All in all, our simulations suggest that: i) Our "Correlation test" is correctly sized in almost every scenario, albeit a little oversized on average, ii) our "friendly-user test" is reasonably well-sized in large samples and iii) the "friendly-user test" becomes oversized in small samples.

Figure 1 shows the kernel density of our correlation-based test. We consider the mid-correlation scenario under the null hypothesis, using 10,000 simulations and sample sizes of T=50, 500, 1000, and 2000. Each density stands for a different sample size. The long-run variances are estimated following Newey and West (1987, 1994). The figure is consistent with the results in Table 1: while small-sample distributions tend to exhibit some distortions and heavier tails, our kernel densities become closer to Gaussian as the sample size increases. In general, for large sample sizes, Figure 1 exhibits well-behaved distributions.

Figure 1: Kernel density of our correlation test with different sample sizes.



Notes: T stands for the sample size. We consider 10,000 Monte Carlo simulations for our "Correlation test" for the mid-correlation scenario under the null hypothesis. Long-run variances are estimated using Newey and West (1987,1994). Source: Author's elaboration.

4.2 Size analysis: one-step-ahead forecasts and heavy-tailed innovations

Here we report simulations using heavy-tailed innovations. Akin to Table 1, we consider three scenarios of correlations under the null hypothesis. In this case, u_t is generated by a $t(5)$ distribution and ε_t by a $t(6)$ distribution; hence, $\sigma_u^2 = 5/3$ and $\sigma_\varepsilon^2 = 3/2$. This is important since we use both parameters to impose the null hypothesis.

For the low-correlation scenario, we set the following parameters: $\phi_y = 0.1, c = 0.1, \phi_x = 0.15, \sigma_\varepsilon^2 = \frac{3}{2}, \sigma_u^2 = \frac{5}{3}$, $Corr(\varepsilon_t, u_t) = 0, \beta_1 = 0.021, \beta_2 = 0.2, \sigma_\omega^2 = 1$, $\sigma_v^2 = 2, \mu_y = 0.1, \mu_x = 0.1$, $\alpha_1 = 0.1$ and $\alpha_2 = 0.1$. With these parameters, we have that $Corr\left(Y_{1,t-1}^f(1), Y_t\right) = Corr\left(Y_{2,t-1}^f(1), Y_t\right) = 0.08$ and therefore, the null hypothesis is imposed.

In the mid-correlation scenario, we set: $\phi_y = 0.4, c = 0.4$, $\phi_x = 0.45$, $\sigma_\varepsilon^2 = \frac{3}{2}$, $\sigma_u^2 = \frac{5}{3}, Corr(\varepsilon_t, u_t) = 0, \beta_1 = 0.2$, $\beta_2 = 0.46$, $\sigma_\omega^2 = 1$, $\sigma_v^2 = 2$, $\mu_y = 0.1$, $\mu_x = 0.1$, $\alpha_1 = 0.1$ and $\alpha_2 = 0.1$. With these parameters, $Corr\left(Y_{1,t-1}^f(1), Y_t\right) = Corr\left(Y_{2,t-1}^f(1), Y_t\right) = 0.455$ and again the null of equality in correlations is satisfied.

Finally, in the high-correlation scenario, we set the following parameters: $\phi_y = 0.65, c = 0.65, \phi_x = 0.65$, $\sigma_\varepsilon^2 = \frac{3}{2}, \sigma_u^2 = \frac{5}{3}, Corr(\varepsilon_t, u_t) = 0, \beta_1 = -0.09, \beta_2 = 0.1, \sigma_\omega^2 = 1$, $\sigma_v^2 = 2, \mu_y = 0.1, \mu_x = 0.1, \alpha_1 = 0.1$ and $\alpha_2 = 0.1$. With these parameters, $Corr\left(Y_{1,t-1}^f(1), Y_t\right) = Corr\left(Y_{2,t-1}^f(1), Y_t\right) = 0.76$ and the null is satisfied.

Tables 2 summarizes our results with fat-tailed innovations. First, akin to Table 1, both tests are reasonably well-sized for large samples ($T \geq 500$). In these cases, the rejections rates of our "Correlation test" range between 9.99% (5.04%) to 11.32% (5.77%) for a nominal size of 10% (5%), with an average size across these six exercises of 10.57%

(5.39%). The "friendly-user test" performs similarly well, with a large sample average size of 10.66% (5.54%). Similar to Table 1, our tests seem to be correctly sized despite the level of correlations under the null hypothesis.

Nevertheless, in contrast to Table 1, some size distortions in small-samples (T≤ 100) become apparent not only for the "friendly-user test", but also for the "Correlation test." The empirical size of the "Correlation test" ranges from 11.48% (5.83%) to 14.51% (8.23%), with a small-sample average size of 12.9% (6.87%). These distortions are even more critical with the "friendly user test": the size ranges from 12.48% (6.90%) to 15.33% (8.90%), with a small-sample average size of 13.93% (7.85%).

Table 2: Size analysis imposing the null hypothesis $H_0: Corr(Y^f_{1,t-1}(1), Y_t) = Corr(Y^f_{2,t-1}(1), Y_t)$ for three different scenarios of correlations under the null. Simulations with heavy-tailed innovations.

(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Correlation scenario	Low-Correlation		Mid-Correlation		High-Correlation	
Nominal Size	*10%*	*5%*	*10%*	*5%*	*10%*	*5%*
			Sample Size=2000			
Correlation test	10.04	5.12	10.37	5.33	9.99	5.04
Friendly-user test	10.14	5.28	10.38	5.37	10.01	5.09
			Sample Size=500			
Correlation test	10.72	5.42	10.99	5.77	11.32	5.67
Friendly-user test	11.00	5.86	11.09	5.88	11.34	5.75
			Sample Size=100			
Correlation test	11.48	5.83	12.69	6.73	12.10	6.47
Friendly-user test	12.79	6.92	12.98	7.05	12.48	6.90
			Sample Size=50			
Correlation test	13.05	6.52	14.51	8.23	13.64	7.44
Friendly-user test	15.33	8.90	15.17	8.85	14.84	8.49

Notes: Each entry represents the percentage of rejections of the null hypothesis of equal correlations using heavy-tailed innovations. Columns (2), (4) and (6) consider a nominal size of 10%, while columns (3), (5) and (7) consider a nominal size of 5%. As both versions of our test are asymptotically normal, we reject the null hypothesis using standard normal critical values. We consider 10,000 Monte Carlo simulations for each exercise. The "friendly user version" of our test requires subtracting the mean of both forecasts and the target variable. Long-run variances are estimated using Newey and West (1987, 1994). Source: Author's elaboration.

4.3 Size analysis: multi-step-ahead forecasts

In this section, we study the size of our tests using multi-step-ahead forecasts. To this end, we impose the null hypothesis at a given forecast horizon h where h = 2, 3, 6 or 12. To impose the null hypothesis, we choose different values for our parameters so that equation (9) is equal to equation (10). From these equations we see that the set of parameters required to satisfy the null hypothesis depends on the horizon h. To keep things as simple as possible, we use as a basic setup the set of parameters used in the mid-correlations scenario of Table 1. The only difference is that now we choose different values of β_2 to impose the null hypothesis at different forecasting horizons h. Table 3 exhibits size results for our tests in multi-step-ahead forecasts. Table 4 is akin to Table 3, but this time we base our parameters on the high-correlations scenario of Table 1. Finally, simulations in Table 5 are based on the mid-correlation scenario of Table 2 with fat-tail innovations.

Table 3: Size analysis imposing the null hypothesis $\mathbf{H_0: Corr\left(Y^f_{1,t-1}(h), Y_{t+h-1}\right) = Corr\left(Y^f_{2,t-1}(h), Y_{t+h-1}\right)}$ for the mid-correlation scenario.

(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Correlation scenario	h=2, $\rho_1 = \rho_2 = 0.467$		h=3, $\rho_1 = \rho_2 = 0.283$		h=6, $\rho_1 = \rho_2 = 0.050$		h=12, $\rho_1 = \rho_2 = 0.002$	
Nominal Size	*10%*	*5%*	*10%*	*5%*	*10%*	*5%*	*10%*	*5%*
Sample Size=2000								
Correlation test	9.30	4.76	9.66	5.06	9.68	4.90	10.54	5.58
Friendly-user test	9.50	4.90	9.86	5.26	9.92	5.20	11.00	5.86
Sample Size=500								
Correlation test	9.78	4.62	10.44	5.28	9.48	5.06	9.82	4.88
Friendly-user test	10.58	5.26	11.20	5.66	10.28	5.60	10.56	5.44
Sample Size=100								
Correlation test	10.80	5.14	10.44	5.14	9.70	4.54	9.82	4.66
Friendly-user test	12.80	7.58	13.04	7.04	12.94	6.56	13.00	6.50
Sample Size=50								
Correlation test	11.84	6.40	10.70	5.16	10.38	4.72	9.48	4.56
Friendly-user test	15.78	9.30	15.04	8.46	15.38	8.14	14.74	7.70

Notes: Each entry represents the percentage of rejections of the null hypothesis of equal correlations. h stands for the forecasting horizon. Columns (2), (4), (6), and (8) consider a nominal size of 10%, while columns (3), (5), (7), and (9) consider a nominal size of 5%. As both versions of our test are asymptotically normal, we reject the null hypothesis using standard normal critical values. We consider 10,000 Monte Carlo simulations for each exercise. The "friendly user" version of our test requires subtracting the mean of both forecasts and the target variable. Long-run variances are estimated using Newey and West (1987, 1994). Source: Author's elaboration.

Table 4: Size analysis imposing the null hypothesis $\mathbf{H_0: Corr\left(Y^f_{1,t-1}(h), Y_{t+h-1}\right) = Corr\left(Y^f_{2,t-1}(h), Y_{t+h-1}\right)}$ for the high-correlation scenario.

(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Correlation scenario	h=2, $\rho_1 = \rho_2 = 0.673$		h=3, $\rho_1 = \rho_2 = 0.531$		h=6, $\rho_1 = \rho_2 = 0.223$		h=12, $\rho_1 = \rho_2 = 0.028$	
Nominal Size	*10%*	*5%*	*10%*	*5%*	*10%*	*5%*	*10%*	*5%*
Sample Size=2000								
Correlation test	9.96	5.22	9.76	5.16	10.62	5.78	10.36	5.36
Friendly-user test	10.01	5.24	9.80	5.16	10.66	5.80	10.38	5.40
Sample Size=500								
Correlation test	10.36	5.26	10.66	5.82	9.98	5.06	9.68	5.32
Friendly-user test	10.54	5.32	10.76	5.98	10.14	5.14	9.74	5.54
Sample Size=100								

Correlation test	12.16	5.90	11.60	6.40	11.26	5.62	10.92	5.40
Friendly-user test	12.82	6.70	12.30	6.86	12.50	6.46	12.00	6.14
				Sample Size=50				
Correlation test	12.64	6.90	12.44	6.66	12.52	6.06	11.60	5.66
Friendly-user test	14.46	8.30	14.44	8.34	14.98	7.92	14.80	7.56

Notes: Each entry represents the percentage of rejections of the null hypothesis of equal correlations. h stands for the forecasting horizon. Columns (2), (4), (6), and (8) consider a nominal size of 10%, while columns (3), (5), (7), and (9) consider a nominal size of 5%. As both versions of our test are asymptotically normal, we reject the null hypothesis using standard normal critical values. We consider 10,000 Monte Carlo simulations for each exercise. The "friendly user" version of our test requires subtracting the mean of both forecasts and the target variable. Long-run variances are estimated using Newey and West (1987, 1994). Source: Author's elaboration.

Table 5: Size analysis imposing the null hypothesis $\mathbf{H_0}$: $\mathbf{Corr}\left(\mathbf{Y^f_{1,t-1}(h)}, \mathbf{Y_{t+h-1}}\right) = \mathbf{Corr}\left(\mathbf{Y^f_{2,t-1}(h)}, \mathbf{Y_{t+h-1}}\right)$ for the mid-correlations scenario. Simulations with heavy-tail innovations.

(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Correlation scenario	h=2, $\rho_1 = \rho_2 = 0.236$		h=3, $\rho_1 = \rho_2 = 0.119$		h=6, $\rho_1 = \rho_2 = 0.014$		h=12, $\rho_1 = \rho_2 = 0.000$	
Nominal Size	*10%*	*5%*	*10%*	*5%*	*10%*	*5%*	*10%*	*5%*
				Sample Size=2000				
Correlation test	10.46	5.56	10.90	5.96	10.16	5.50	9.98	5.34
Friendly-user test	10.48	5.58	10.90	5.96	10.24	5.58	10.04	5.38
				Sample Size=500				
Correlation test	11.06	5.54	10.40	5.58	10.56	5.42	10.44	5.08
Friendly-user test	11.06	5.54	10.48	5.68	10.68	5.58	10.46	5.24
				Sample Size=100				
Correlation test	12.38	6.88	12.22	6.76	12.14	6.90	11.60	5.84
Friendly-user test	12.80	7.16	12.64	7.06	12.48	7.04	12.22	6.34
				Sample Size=50				
Correlation test	14.72	7.92	14.50	7.74	13.12	6.88	14.20	7.64
Friendly-user test	15.36	8.86	15.36	8.30	14.16	7.90	15.58	8.46

Notes: Each entry represents the percentage of rejections of the null hypothesis of equal correlations. h stands for the forecasting horizon. Columns (2), (4), (6), and (8) consider a nominal size of 10%, while columns (3), (5), (7), and (9) consider a nominal size of 5%. As both versions of our test are asymptotically normal, we reject the null hypothesis using standard normal critical values. We consider 10,000 Monte Carlo simulations for each exercise. The "friendly user" version of our test requires subtracting the mean of both forecasts and the target variable. Long-run variances are estimated using Newey and West (1987, 1994). Source: Author's elaboration.

There are some features in Tables 3 through 5 worth mentioning. First, both tests perform remarkably well in large samples (T≥500): considering the three tables, the empirical size of the "Correlation test" ranges from 9.30% (4.62%) to 11.02% (5.96%). The "friendly user" version of our test reports almost the same size as the main "Correlation test" in large samples. These results are interesting since our tables consider different levels of correlations under the null and different forecasting horizons. Second, contrary to the expected, we do not see a clear pattern of size distortions with the forecasting horizon h. Third, similar to our previous simulations, the "friendly-user test" becomes significantly oversized in small samples (T=50). Fourth, the "Correlation test" is reasonably well-sized even in small samples, except for the case with fat-tail innovations in Table 5. For instance, the empirical size of our test in Table 3 ranges from 9.48% (4.56%) through 11.84% (6.40%). In contrast, Table 5 exhibits rejections ranging from 13.12% (6.88%) to 14.72% (7.92%).

4.4 Power analysis one-step-ahead

Here we explore the power of our tests using one-step-ahead forecasts. To this end, we consider as a start point the following parameters of our mid-correlations scenario: $\phi_y = 0.5, c = 0.7, \phi_x = 0.4, \sigma_u^2 = 1.5, \sigma_\varepsilon^2 = 1,\ Corr(\varepsilon_t, u_t) = 0.5, \beta_1 = 0.6, \beta_2 = 0.77, \sigma_\omega^2 = 2,\ \ \sigma_v^2 = 2.5, \mu_y = 0.3, \mu_x = 0.4, \alpha_1 = 0.1$ and $\alpha_2 = 0.2$. Recall that, under these parameters, we impose the null hypothesis at $Corr\left(Y_{1,t-1}^f(1), Y_t\right) = Corr\left(Y_{2,t-1}^f(1), Y_t\right) = 0.687$. To impose the alternative hypothesis, we shrink the coefficient β_2 to four different values: 0.562, 0.369, 0.269, and 0.12. With these new values of β_2, the differences in correlations become $Corr\left(Y_{1,t-1}^f(1), Y_t\right) - Corr\left(Y_{2,t-1}^f(1), Y_t\right) = \rho_1 - \rho_2 = 0.035, 0.077, 0.102$ and 0.15 respectively. We consider four different sample sizes: 50, 100, 500, 1000, and 2000. Finally, we run 10,000 Monte Carlo replications with nominal sizes of 10% and 5%.

Table 6: Power analysis imposing the alternative hypothesis for a mid-correlations scenario.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Under the alternative hypothesis:	$\rho_1 - \rho_2 = 0.035$		$\rho_1 - \rho_2 = 0.077$		$\rho_1 - \rho_2 = 0.102$		$\rho_1 - \rho_2 = 0.150$	
Nominal Size	10%	5%	10%	5%	10%	5%	10%	5%
			Sample Size = 2000					
Correlation test	83.73	74.39	100.00	99.97	100.00	100.00	100.00	100.00
Friendly-user test	84.12	75.00	100.00	99.97	100.00	100.00	100.00	100.00
			Sample Size = 1000					
Correlation test	58.04	45.19	98.46	96.62	99.92	99.80	100.00	100.00
Friendly-user test	58.62	46.62	98.53	96.82	99.94	99.81	100.00	100.00
			Sample Size = 500					
Correlation test	36.97	25.52	84.64	75.68	96.19	92.73	99.91	99.67
Friendly-user test	37.98	27.02	85.28	76.98	96.40	93.11	99.91	99.67
			Sample Size = 100					
Correlation test	16.17	8.85	31.9	19.91	45.06	31.54	64.87	50.38
Friendly-user test	18.96	11.62	35.22	24.57	48.55	36.97	67.70	55.61
			Sample Size = 50					
Correlation test	13.85	6.72	21.85	12.04	27.19	16.09	40.14	25.97
Friendly-user test	18.55	10.86	27.28	17.75	32.53	22.27	46.52	33.86

Notes: Each entry represents the percentage of rejections of the null hypothesis of equal correlations. The alternative hypothesis is set at four different levels: $\rho_1 - \rho_2 = 0.035, 0.077, 0.102$ and 0.15 respectively. Columns (2), (4), (6) and (8) consider a nominal size of 10%, while columns (3), (5), (7) and (9) consider a nominal size of 5%. As both versions of our test are asymptotically normal, we reject the null hypothesis using standard normal critical values. We consider 10,000 Monte Carlo simulations for each exercise. The "friendly user" version of our test requires subtracting the mean of both forecasts and the target variable. Long-run variances are estimated using Newey and West (1987, 1994). Source: Author's elaboration.

Table 6 exhibits power results. First, for large samples (T ≥ 500) and $\rho_1 - \rho_2 \geq 0.077$, our test has an outstanding ability to detect differences in correlations. For instance, columns (4) through (9) show rejections rates ranging from 75.68% to 100%. Notably, for a sample size of 1000, rejections range between 96.62% and 100%. These results are particularly interesting since the largest difference considered in these simulations is just $\rho_1 - \rho_2 = 0.15$; hence, in large samples, even small differences in correlations are frequently detected by our test.

Second, and naturally, the power of our test drops with the sample size (particularly for T=50). For instance, the outstanding 100% rejection in column (8) for large samples reduces to (about) 40% in small samples. Of course, these results are even worst for smaller differences in correlations under the alternative: For $\rho_1 - \rho_2 = 0.035$, the rejection rate in small samples is just above nominal size. Finally, we do not observe significant differences between both versions of our test, although the "friendly-user" version tends to be slightly more powerful, most likely due to the bigger size distortions associated with it. All in all, our test has a reasonable power with large samples, or with sufficiently large differences in correlations under the alternative.

4.5 Power analysis multi-steps-ahead

Here we explore the power of our tests using multi-step-ahead forecasts. We consider four different forecasting horizons h=2, 3, 6, and 12. Based on the high-correlation scenario, we consider as a start point the following parameters: $\phi_y = 0.8, c = 0.8, \phi_x = 0.8, \sigma_u^2 = 1, \sigma_\varepsilon^2 = 1,$ $Corr(\varepsilon_t, u_t) = 0, \beta_1 = 0.6, \beta_2 = 0.807, \sigma_\omega^2 = 1, \sigma_v^2 = 1.3, \mu_y = 0.1, \mu_x = 0.1, \alpha_1 = 0.1$ and $\alpha_2 = 0.1$. Under these parameters, we impose the null hypothesis at $Corr\left(Y_{1,t-1}^f(1), Y_t\right) =$

$Corr\left(Y_{2,t-1}^{f}(1), Y_t\right) = 0.96$. We impose the alternative hypothesis $Corr\left(Y_{1,t-1}^{f}(h), Y_{t+h-1}\right) - Corr\left(Y_{2,t-1}^{f}(h), Y_{t+h-1}\right) = 0.1, \forall h \in \{2,3,6,12\}$. To impose the alternative hypothesis for each forecasting horizon, we change the coefficient β_2 to -0.427 (h=2), -0.326 (h=3), -0.324 (h=6), and -0.889 (h=12). We consider four different sample sizes: 50, 100, 500, 1000, and 2000. Finally, we run 10,000 Monte Carlo replications with nominal sizes of 10% and 5%.

Each entry in Table 7 reports the rate of rejections of the null hypothesis of equal correlations. Two features of this table are worth mentioning. First, consistent with Table 6, our test has an outstanding ability to recognize differences in correlations for large samples: For T≥ 1,000, the rate of rejection of our test is above 99% for all the forecasting horizons and nominal sizes. Second, we observe a deterioration of power for longer forecasting horizons in small samples (T≤ 100): While the rate of rejection of our test ranges from 82.12% to 99.98% for h=2 steps-ahead-forecasts, it dramatically drops to range between 10.81 to 32.24 in our h=12 steps-ahead exercises. However, we emphasize that this pattern of power deterioration with the forecasting horizon disappears rapidly with a larger sample size.

Table 7: Power analysis imposing the alternative hypothesis $Corr\left(Y_{1,t-1}^{f}(h), Y_{t+h-1}\right) - Corr\left(Y_{2,t-1}^{f}(h), Y_{t+h-1}\right) = 0.1$ in multi-step-ahead forecasts

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Under the alternative hypothesis:	h=2, $\rho_1 - \rho_2 = 0.1$		h=3, $\rho_1 - \rho_2 = 0.1$		h=6, $\rho_1 - \rho_2 = 0.1$		h=12, $\rho_1 - \rho_2 = 0.1$	
Nominal Size	10%	5%	10%	5%	10%	5%	10%	5%
Sample Size = 2000								
Correlation test	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Friendly-user test	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Sample Size = 1000								
Correlation test	100.00	100.00	100.00	100.00	100.00	100.00	99.65	99.04
Friendly-user test	100.00	100.00	100.00	100.00	100.00	100.00	99.66	99.04
Sample Size = 500								
Correlation test	100.00	100.00	100.00	100.00	100.00	100.00	90.96	84.44
Friendly-user test	100.00	100.00	100.00	100.00	100.00	100.00	90.94	84.52
Sample Size = 100								
Correlation test	99.88	99.25	98.89	96.84	83.14	69.79	30.38	18.68
Friendly-user test	99.98	99.85	99.3	98.24	84.85	73.57	32.24	20.13
Sample Size = 50								
Correlation test	92.29	82.12	83.16	68.22	50.00	31.73	19.18	10.81
Friendly-user test	96.05	96.05	88.44	77.86	56.32	38.82	26.77	15.90

Notes: Each entry represents the percentage of rejections of the null hypothesis of equal correlations. The alternative hypothesis is set at $\rho_1 - \rho_2 = 0.1$. h stands for the forecasting horizon. Columns (2), (4), (6) and (8) consider a nominal size of 10%, while columns (3), (5), (7) and (9) consider a nominal size of 5%. As both versions of our test are asymptotically normal, we reject the null hypothesis using standard normal critical values. We consider 10,000 Monte Carlo simulations for each exercise. The "friendly user" version of our test requires subtracting the mean of both forecasts and the target variable. Long-run variances are estimated using Newey and West (1987, 1994). Source: Author's elaboration.

5. Empirical illustration

Here we illustrate the relevance of our test when forecasting headline inflation. To this end, we consider two competing forecasts for monthly consumer price indices (CPI) in 35 OECD economies (see Table 8). Our first forecasting model considers an international factor as a predictor of local inflation. As pointed out by Pincheira and Gatty (2016): "*a few relatively recent articles report an important pass-through from some measures of industrialized international inflation to local inflation. In particular, Ciccareli and Mojon (2010) and West (2008) find that local inflation in OECD countries is importantly driven by a common inflation factor*." Pincheira and Gatty (2016) page 2. See Medel, Pedersen and Pincheira (2016) for a comprehensive analysis of 31 OECD economies. Other interesting articles supporting this view are Duncan and Martínez-García (2015), Kabukcuoglu and Martínez-García (2018), Morales-Arias and Moura (2013), Hakkio (2009), Pincheira and Gatty (2016), Medel, Pedersen and Pincheira (2016) and Pincheira (2022).

Our second forecasting model considers core inflation as a predictor of headline inflation. As commented by Pincheira, Selaive and Nolazco (2019): "*[…] the 'core predicts headline' argument is fairly popular. In a context in which inflation is not easy to forecast (Stock and Watson, 2008) the idea that core inflation may be a useful predictor in principle is very appealing, especially for central banks that are responsible for maintaining overall price stability and need to know where inflation is heading.*" Pincheira, Selaive and Nolazco (2019), page 1060.

Following Robalo, Duarte and Morais (2003), we consider core inflation as CPI excluding "food" and "energy" components. The rationale for using core inflation is that "food" and "energy" tend to be highly volatile components and removing them may improve the predictive performance of many models. While there is some evidence supporting the predictive ability of core inflation (e.g., Bermingham (2007) and Song (2005), among many others), it is our reading of the literature that the usefulness of core inflation as a predictor of headline inflation is not equal across countries, and more generally, is still an open debate (see Bullard (2011), Le Bihan and Sédillot (2000) and Freeman (1998) for detractors). See Pincheira, Selaive and Nolazco (2019) for an interesting discussion.

Our data is collected from the OECD database from January 2000 through December 2021 at the monthly frequency for all the economies under analysis[8]. Let π_t be the year-on-year headline inflation rate for a given economy at month "t", π_t^{Core} the corresponding year-on-year core inflation, and π_t^{Int} the OECD year-on-year headline inflation (as a proxy of the international factor). Our competing forecasting models are:

$$\pi_{t+1} - \pi_t = c_1 + \rho_1(\pi_t - \pi_{t-1}) + \beta_1(\pi_t^{Int} - \pi_{t-1}^{Int}) + \varepsilon_{1,t+1} \qquad (Model\ 1)$$

$$\pi_{t+1} - \pi_t = c_2 + \rho_2(\pi_t - \pi_{t-1}) + \beta_2\left(\pi_t^{Core} - \pi_{t-1}^{Core}\right) + \varepsilon_{2,t+1} \qquad (Model\ 2)$$

Where $\varepsilon_{1,t+1}$ and $\varepsilon_{2,t+1}$ are error terms.

To conduct our out-of-sample analysis, we split our sample into two parts: an initial estimation window of size R and a prediction window of size P (note that P+R=T, where T is the total number of observations and P is the number of one-step-ahead forecasts). We update our OLS estimators using rolling windows of R=50 observations.

[8] https://stats.oecd.org/

Let $\Delta Correlations = Correlation_{Model\ 2} - Correlation_{Model\ 1}$, where $Correlation_{Model\ i}$ is simply the correlation between our forecast constructed with model $i \in \{1,2\}$ and the target variable. Let $Ratio\ MSPE = \frac{MSPE_{Model\ 1}}{MSPE_{Model\ 2}}$, where $MSPE_{Model\ i}$ is simply the MSPE of model $i \in \{1,2\}$. In order to evaluate the null hypothesis of equal MSPE, we conduct the Diebold and Mariano (1995) and West (1996) (DMW) test, using a HAC estimator following Newey and West (1987, 1994).

Table 8 next reports $\Delta Correlations$ and $Ratio\ MSPE$ for each country. Bold entries highlight rejections of the null hypothesis. Three features in Table 8 are worth mentioning. First, the DMW test rejects the null hypothesis in 6 out of 35 exercises at usual significance levels. Notably, these 6 exercises suggest that Model 1 outperforms Model 2 in terms of MSPE. In sharp contrast, our test rejects the null hypothesis of equal correlations in 14 out of 35 exercises. Among these 14 rejections, 86% of them suggest that Model 1 outperforms Model 2 in terms of correlations. In sum, our test is able to detect differences between the two forecasts, even when the differences in MSPE are not encouraging. Second, in 60% of our exercises, we do not observe significant differences for MSPE nor correlations; in this sense, there are several countries in which we do not find significant differences between predicting with core inflation or with the OECD international factor. Finally, for the 6 cases in which both tests reject the null hypothesis, our test tends to reject at a tighter significance level relative to DMW: half of the times both tests reject the null at the same significance level, and half of the times our test rejects at a tighter significance level. All in all, our test tends to reject the null hypothesis more frequently than DMW, and in many cases at a tighter significance level.

Table 8. Forecasting headline inflation: Comparing a test of equal MSPE with our correlation-based test.

	Austria	Belgium	Canada	Chile	Colombia	Costa Rica	Czech Republic
Δ Correlations	**-0.171****	0.009	**0.153***	**-0.082****	**-0.058****	**-0.231*****	**-0.133***
Ratio MSPE	**0.884***	0.985	1.082	**0.905****	**0.929****	**0.808****	0.965
	Denmark	Estonia	Finland	France	Germany	Greece	Hungary
Δ Correlations	0.064	**-0.130***	**-0.067***	0.023	-0.013	-0.087	0.000
Ratio MSPE	1.027	**0.901***	0.957	1.017	1.003	0.956	1.015
	Iceland	Ireland	Israel	Italy	Japan	Korea	Latvia
Δ Correlations	-0.006	-0.007	**-0.137***	-0.044	-0.065	-0.136	0.019
Ratio MSPE	1.014	0.990	0.891	0.971	0.958	0.921	1.028
	Lithuania	Luxembourg	Mexico	Netherlands	Norway	Poland	Portugal
Δ Correlations	0.000	-0.102	0.050	**0.160***	-0.056	**-0.070****	-0.051
Ratio MSPE	1.001	0.922	1.057	1.049	0.977	0.948	0.964
	Slovak Republic	Slovenia	Spain	Sweden	Switzerland	Turkey	UK
Δ Correlations	-0.078	0.091	0.031	**-0.132***	**-0.374*****	**-0.045***	-0.113
Ratio MSPE	0.961	1.078	1.018	0.938	**0.812****	0.953	0.910

Notes: $\Delta Correlations$ is defined as $Correlation_{Model\ 2} - Correlation_{Model\ 1}$, where $Correlation_{Model\ i}$ is the correlation between our forecast constructed with model $i \in \{1,2\}$ and the target variable. $Ratio\ MSPE$ is defined as $\frac{MSPE_{Model\ 1}}{MSPE_{Model\ 2}}$, where $MSPE_{Model\ i}$ is the MSPE of model $i \in$

{1,2}. We consider the Diebold and Mariano (1995) and West (1996) test to evaluate the null hypothesis of equal MSPE, and our test to evaluate the null of equal correlations. *p<0.1, **p<0.05,***p<0.01.

6. Concluding remarks

In this paper, we propose a correlation-based test for the evaluation of two competing forecasts. Under the null hypothesis of equal correlations with the target variable, we derive the asymptotic distribution of our test using the delta method. A recent paper by Pincheira and Hardy (2021) (henceforth PH) demonstrates that the null hypothesis of equal correlations is not (necessarily) equivalent to a null hypothesis of equal MSPE. In particular, they show that the forecast displaying the lowest MSPE may also exhibit the lowest correlation with the target variable. They label this result as "The MSPE Paradox." While this is an interesting observation, they do not provide formal procedures to test the null hypothesis of equal correlations. This paper fills this gap.

We provide a correlation-based test and a simpler "friendly-user" version of it. Monte Carlo simulations suggest that both tests are reasonably well-sized in large samples. Our tests seem to be correctly sized even considering different levels of correlations under the null hypothesis and different forecasting horizons. In small samples, however, some distortions become apparent for the "friendly-user" test.

Our empirical application, in which we compare two forecasts for headline inflation, clearly illustrates the benefits of our correlation-based test: sometimes, traditional tests comparing MSPE cannot detect differences between two competing forecasts, despite one of them having significantly higher correlations. In this sense, our test can be viewed as a complement to the traditional tests of Diebold and Mariano (1995) and West (1996) (DMW). Our illustration reveals that the DMW test of equal MSPE rejects the null hypothesis only for 17% of our exercises; in sharp contrast, our correlation-based test rejects the null hypothesis in 40% of our exercises. Notably, our test not only detects differences more frequently but also at equal or tighter significance levels.

There are two interesting avenues for future research. First, even though our test is useful in many applications, it relies on the assumption that the correlation of both forecasts exists. One important caveat of this assumption is that it rules out comparisons with a forecast with zero variance. One leading example is the zero-forecast, which is a common benchmark in the literature. Hence, an extension of our test to the zero variance case is warranted. Second, our approach considers forecasts as primitives. Consequently, we are silent about the effects of parameter uncertainty when models are used to generate the forecasts under analysis. Another natural extension of our paper could incorporate these effects along the lines of West (1996).

7. References

1. Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish journal of emergency medicine*, *18*(3), 91-93.
2. Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, 817-858.
3. Ang, A., Bekaert, G., and Wei, M. (2007) Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, Volume 54, issue 4, 1163-1212.

4. Bermingham, C. (2007). How Useful is Core Inflation for Forecasting Headline Inflation?. *Economic & Social Review*, *38*(3).
5. Bentancor, A., and Pincheira, P. (2010) Predicción de errores de proyección de inflación en Chile. *El Trimestre Económico*, Vol (0) 305, 129-154. (In Spanish).
6. Bullard, J. (2011). Measuring inflation: the core is rotten. *Federal Reserve Bank of St. Louis Review*, *93*(4), 223-233.
7. Busetti, F., and Marcucci, J. (2013). Comparing forecast accuracy: a Monte Carlo investigation. *International Journal of Forecasting*, *29*(1), 13-27.
8. Ciccarelli, M., and Mojon, B. (2010). Global inflation*. The Review of Economics and Statistics*, *92(3)*, 524-535.
9. Clark, T., and McCracken, M. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, *105*(1), 85-110.
10. Clark, T., and McCracken, M. (2005): "Evaluating Direct Multistep Forecasts". *Econometric Reviews* 24(4): 369-404.
11. Clark, T., and McCracken, M. (2013). Advances in forecast evaluation. *Handbook of Economic Forecasting*, *2*, 1107-1201.
12. Clark, T., and West, K. (2006): "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis". *Journal of Econometrics* **135**: 155-186.
13. Clark, T., and West, K. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291-311.
14. Diebold, F., and Mariano, R. (1995): "Comparing Predictive Accuracy". *Journal of Business & Economic Statistics* 13: 253-263.
15. Duncan, R., and Martínez-García, E. (2015). Forecasting local inflation with global inflation: When economic theory meets the facts. *Globalization and Monetary Policy Institute Working Paper*, (235).
16. Elliott, G., and Timmermann, A. (2013). *Handbook of Economic Forecasting*. Elsevier.
17. Freeman, D. G. (1998). Do core inflation measures help forecast inflation?. *Economics Letters*, *58*(2), 143-147.
18. Giacomini, R., and White,H. (2006): "Tests of Conditional Predictive Ability". *Econometrica* 74(6): 1545-1578.
19. Granger, C., and Machina, J. (2006). Forecasting and decision theory. *Handbook of Economic Forecasting*, *1*, 81-98.
20. Granger, C. (1999). Empirical modeling in economics: Specification and evaluation. *Cambridge University Press.*
21. Granger, C., and Newbold, P. (1986). Forecasting Economic Time Series. *Elsevier Monographs*.
22. Hakkio, C. (2009). Global inflation dynamics. *Federal Reserve Bank of Kansas City Working Paper*, (09-01).
23. Hansen, B. (2019). Econometrics. *University of Wisconsin. www.ssc.wisc.edu/~bhansen.*
24. 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*, 131-151.
25. Joutz, F. and Stekler, H., (2000). An Evaluation of the Prediction of the Federal Reserve. *International Journal of Forecasting*. 16: 17–38.
26. Kabukçuoğlu, A., and Martínez-García, E. (2018). Inflation as a global phenomenon—Some implications for inflation modeling and forecasting. *Journal of Economic Dynamics and Control*, *87*, 46-73.
27. Le Bihan, H., and Sédillot, F. (2000). Do core inflation measures help forecast inflation?: Out-of-sample evidence from French data. *Economics Letters*, *69*(3), 261-266.
28. McCracken M. (2007): "Asymptotics for Out of Sample Tests of Granger Causality". *Journal of Econometrics* 140: 719-752.

29. Medel, C., Pedersen, M., and Pincheira, P. (2016). The elusive predictive ability of global inflation. *International Finance*, *19*(2), 120-146.
30. Mincer, J. A., and Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3-46). NBER.
31. Morales-Arias, L., and Moura, G. V. (2013). A conditionally heteroskedastic global inflation model. *Journal of Economic Studies*.
32. Newey, W., and West, K. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777-787.
33. Newey, W., and West, K. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, *61*(4), 631-653.
34. Nordhaus, W. (1987). Forecasting efficiency: concepts and applications. *The Review of Economics and Statistics*, Vol LXIX, N°4.
35. Patton, A., and Timmermann, A. (2012). Forecast Rationality Tests Based on Multi-Horizon Bounds ". *Journal of Business and Economic Statistics* 30(1): 1–17.
36. Pincheira, P. (2022). A Power Booster Factor for Out-of-Sample Tests of Predictability. *Economia*, *forthcoming*.
37. Pincheira, P. (2010). A Real time evaluation of the Central Bank of Chile GDP growth forecasts. *Money Affairs*, CEMLA, vol. 0(1), 37-73.
38. Pincheira, P. (2012). A joint test of superior predictive ability for Chilean inflation forecasts. *Journal Economía Chilena* (The Chilean Economy), Central Bank of Chile, vol. 15(3), 04-39, (In Spanish).
39. Pincheira, P., and Álvarez, R. (2009). Evaluation of short run inflation forecasts and forecasters in Chile," *Money Affairs*, CEMLA, vol. 0(2), 159-180.
40. Pincheira, P. and Fernández, N. (2011). Corrección de algunos errores sistemáticos de predicción de inflación. *Monetaria*, CEMLA, vol. 0(1), 37-61. (In Spanish).
41. Pincheira, P., and Gatty, A. (2016). Forecasting Chilean inflation with international factors. *Empirical Economics*, *51*(3), 981-1010.
42. Pincheira, P., and Hardy, N. (2021). The Mean Squared Prediction Error Paradox (No. 107403). *University Library of Munich, Germany.*
43. Pincheira, P., Hardy, N., and Muñoz, F. (2021). "Go Wild for a While!": A New Test for Forecast Evaluation in Nested Models. *Mathematics*, *9*(18), 2254.
44. Pincheira, P., Selaive, J., and Nolazco, J. (2019). Forecasting inflation in Latin America with core measures. *International Journal of Forecasting*, *35*(3), 1060-1071.
45. Robalo, C., Duarte, P., and Morais, L. (2003). "Evaluating core inflation indicators." *Economic Modelling* 20, 765-775.
46. Samuel, M., and Okey, L. (2015). The relevance and significance of correlation in social science research. *International Journal of Sociology and Anthropology Research*, *1*(3), 22-28.
47. Song, L. (2005). Do underlying measures of inflation outperform headline rates? Evidence from Australian data. *Applied Economics*, *37*(3), 339-345.
48. West K. (1996):" Asymptotic Inference About Predictive Ability". *Econometrica* 64: 1067-1084.
49. West K. (2006): "Forecast Evaluation" *Handbook of Economic Forecasting*, Elsevier.
50. White, H (2000). *Asymptotic theory for econometricians*. Academic press.

8. Appendices

Appendix A1. Our decomposition of MSPE assuming Var(X)=Var(Z), E(X)=0 and $E[X^2] > 0$

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

$$= (EX^2 - EZ^2) - 2(EYX - EYZ)$$

$$= (EX^2 - EZ^2) - 2\{Cov(Y,X) - Cov(Y,Z)\}$$

$$= [MSF_X - MSF_Z] - 2\{Cov(Y,X) - Cov(Y,Z)\}$$

$$= [MSF_X - MSF_Z] - 2\sqrt{Var(Y)}\left\{Corr(Y,X)\sqrt{Var(X)} - Corr(Y,Z)\sqrt{Var(Z)}\right\}$$

$$= [MSF_X - MSF_Z] - 2\sqrt{Var(Y)}\sqrt{Var(X)}\{Corr(Y,X) - Corr(Y,Z)\}$$

$$[MSF_X - MSF_Z] - 2\sqrt{Var(Y)}\sqrt{MSF_X - (EX)^2}\{Corr(Y,X) - Corr(Y,Z)\}$$

$$[MSF_X - MSF_Z] - 2\sqrt{Var(Y)}\sqrt{MSF_X}\{Corr(Y,X) - Corr(Y,Z)\} \blacksquare$$

Appendix A2. Our general decomposition of MSPE (dropping our previous assumptions).

$$\Delta MSPE = MSF_X - MSF_Z - 2\left\{Corr(Y,X)\sqrt{V(Y)V(X)} - Corr(Y,Z)\sqrt{V(Y)V(Z)} + EY(EX - EZ)\right\}$$

$$= MSF_X - MSF_Z - 2\sqrt{V(Y)}\left\{Corr(Y,X)\sqrt{V(X)} - Corr(Y,Z)\sqrt{V(Z)}\right\} - 2\{EY(EX - EZ)\}$$

$$= MSF_X - MSF_Z - 2\sqrt{V(Y)}\left\{Corr(Y,X)\sqrt{MSF_x - \mu_x^2} - Corr(Y,Z)\sqrt{MSF_z - \mu_z^2}\right\} - 2\{EY(EX - EZ)\} \blacksquare$$

Appendix A3. Additional simulations

Recall from Section 4 that our VAR(1) DGP is given by

$$Y_t = \mu_y + \phi_y Y_{t-1} + cX_{t-1} + \varepsilon_t \quad (A1)$$

$$X_t = \mu_x + \phi_x X_{t-1} + u_t \quad (A2)$$

$$\begin{pmatrix}\varepsilon_t \\ u_t\end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho_{\varepsilon,u}\sigma_u \\ \rho_{\varepsilon,u}\sigma_u & \sigma_u^2 \end{pmatrix}\right)$$

Akin to Section 4, let us consider $Z_{1,t}$ and $Z_{2,t}$ as two proxies of the unobservable variable X_t, generated as

$$Z_{1,t} = \alpha_1 + \beta_1 X_t + \omega_t \quad (A3)$$

$$Z_{2,t} = \alpha_2 + \beta_2 X_t + v_t \quad (A4)$$

where the pair the pair $(\omega_t, v_t)'$ is a Gaussian white noise vector that is totally independent from pair $(\varepsilon_t, u_t)'$.

Now let us consider our FGP with the following two competing one-step-ahead forecasts for Y_t:

$$Y_{1,t-1}^{f}(1) = \mu_1 + \phi_1 Y_{t-1} + c_1 Z_{1,t-1} \quad (A5)$$

$$Y_{2,t-1}^{f}(1) = \mu_2 + \phi_2 Y_{t-1} + c_2 Z_{2,t-1} \quad (A6)$$

In contrast to Section 4, these FGP (A5) and (A6) allow μ_1 and μ_2 to be different from μ_y, ϕ_1 and ϕ_2 to be different from ϕ_y, and c_1 and c_2 to be different from c.

To assess the empirical size of our test with this new parameterization, we impose the null hypothesis of equal correlation with the target variable $H_0: Corr\left(Y_{1,t-1}^{f}(h), Y_{t+h-1}\right) = Corr\left(Y_{2,t-1}^{f}(h), Y_{t+h-1}\right)$ by choosing specific values of the following parameters: $\phi_1, \phi_2, c_1, c_2, \phi_y, c, \phi_x, \rho_{\varepsilon,u}, \sigma_\varepsilon, \sigma_u, \beta_1, \beta_2, V(\omega_t)$ and $V(v_t)$. To that end, we make use of the following straightforward results that stem from the definition of our DGP and FGP:

Since X_t is an AR(1) process, it follows from equation (A2) that

$$V(X_t) = \frac{\sigma_u^2}{1-\phi_x^2}$$

From equations (A1) and (A2)

$$Cov(Y_t, X_t) = \frac{\phi_x c V(X_t) + Cov(u_t, \varepsilon_t)}{1-\phi_x \phi_y}$$

$$V(Y_t) = \frac{c^2 V(X_t) + 2\phi_y c Cov(Y_t, X_t) + \sigma_\varepsilon^2}{1-\phi_y^2}$$

Taking (A1) together with (A3) and (A4)

$$Cov(Z_{1t}, Y_t) = \beta_1 Cov(X_t, Y_t)$$

$$Cov(Z_{2t}, Y_t) = \beta_2 Cov(X_t, Y_t)$$

Taking (A2) together with (A3) and (A4)

$$Cov(Z_{1t}, X_t) = \beta_1 V(X_t)$$

$$Cov(Z_{2t}, X_t) = \beta_2 V(X_t)$$

The variances of equations (A5) and (A6) are given by

$$V\left(Y_{1,t-1}^{f}(1)\right) = \phi_1^2 V(Y_t) + c_1^2\left(\beta_1^2 V(X_t) + V(\omega_t)\right) + 2\phi_1 c_1 Cov\left(Z_{1,t}, Y_t\right)$$

$$V\left(Y_{2,t-1}^{f}(1)\right) = \phi_2^2 V(Y_t) + c_2^2\left(\beta_2^2 V(X_t) + V(v_t)\right) + 2\phi_2 c_2 Cov\left(Z_{2,t}, Y_t\right)$$

Hence, using equation (A1) together with equations (A5) and (A6)

$$Cov\left(Y_{1,t-1}^{f}(1), Y_t\right) = \phi_y\phi_1 V(Y_t) + c_1\phi_y Cov\left(Y_t, Z_{1,t}\right) + c\phi_1 Cov(X_t, Y_t) + cc_1 Cov\left(X_t, Z_{1,t}\right)$$

$$Cov\left(Y_{2,t-1}^{f}(1), Y_t\right) = \phi_y\phi_2 V(Y_t) + c_2\phi_y Cov\left(Y_t, Z_{2,t}\right) + c\phi_2 Cov(X_t, Y_t) + cc_2 Cov\left(X_t, Z_{2,t}\right)$$

$$Corr\left(Y_{1,t-1}^{f}(1), Y_t\right) =$$

$$\frac{\phi_y\phi_1 \dfrac{c^2\dfrac{\sigma_u^2}{1-\phi_x^2} + 2\phi_y c\dfrac{\phi_x c\dfrac{\sigma_u^2}{1-\phi_x^2} + \rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y} + \sigma_\varepsilon^2}{1-\phi_y^2} + \dfrac{\phi_x c\dfrac{\sigma_u^2}{1-\phi_x^2} + \rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y}(c_1\phi_y\beta_1 + c\phi_1) + cc_1\beta_1\dfrac{\sigma_u^2}{1-\phi_x^2}}{\sqrt{\left(\dfrac{c^2\dfrac{\sigma_u^2}{1-\phi_x^2} + 2\phi_y c\dfrac{\phi_x c\dfrac{\sigma_u^2}{1-\phi_x^2} + \rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y} + \sigma_\varepsilon^2}{1-\phi_y^2}\right)\left(\phi_1^2\dfrac{c^2\dfrac{\sigma_u^2}{1-\phi_x^2} + 2\phi_y c\dfrac{\phi_x c\dfrac{\sigma_u^2}{1-\phi_x^2} + \rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y} + \sigma_\varepsilon^2}{1-\phi_y^2} + c_1^2\left(\beta_1^2\dfrac{\sigma_u^2}{1-\phi_x^2} + V(\omega_t)\right) + 2\phi_1 c_1\beta_1\dfrac{\phi_x c\dfrac{\sigma_u^2}{1-\phi_x^2} + \rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y}\right)}} \quad (A7)$$

$$Corr\left(Y_{2,t-1}^{f}(1), Y_t\right) =$$

$$\frac{\phi_y\phi_2 \dfrac{c^2\dfrac{\sigma_u^2}{1-\phi_x^2} + 2\phi_y c\dfrac{\phi_x c\dfrac{\sigma_u^2}{1-\phi_x^2} + \rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y} + \sigma_\varepsilon^2}{1-\phi_y^2} + \dfrac{\phi_x c\dfrac{\sigma_u^2}{1-\phi_x^2} + \rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y}(c_2\phi_y\beta_2 + c\phi_2) + cc_2\beta_2\dfrac{\sigma_u^2}{1-\phi_x^2}}{\sqrt{\left(\dfrac{c^2\dfrac{\sigma_u^2}{1-\phi_x^2} + 2\phi_y c\dfrac{\phi_x c\dfrac{\sigma_u^2}{1-\phi_x^2} + \rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y} + \sigma_\varepsilon^2}{1-\phi_y^2}\right)\left(\phi_2^2\dfrac{c^2\dfrac{\sigma_u^2}{1-\phi_x^2} + 2\phi_y c\dfrac{\phi_x c\dfrac{\sigma_u^2}{1-\phi_x^2} + \rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y} + \sigma_\varepsilon^2}{1-\phi_y^2} + c_2^2\left(\beta_2^2\dfrac{\sigma_u^2}{1-\phi_x^2} + V(v_t)\right) + 2\phi_2 c_2\beta_2\dfrac{\phi_x c\dfrac{\sigma_u^2}{1-\phi_x^2} + \rho_{\varepsilon,u}\sigma_\varepsilon\sigma_u}{1-\phi_x\phi_y}\right)}} \quad (A8)$$

Now we impose the null hypothesis at different levels of correlations. The idea is to choose specific values for the parameters $\phi_1, \phi_2, c_1, c_2, \phi_y, c, \phi_x, \rho_{\varepsilon,u}, \sigma_\varepsilon, \sigma_u, \beta_1, \beta_2, V(\omega_t)$ and $V(v_t)$, so that equation (A7) is equal to equation (A8). Akin to Section 4, we consider three scenarios: 1) A low-correlation scenario ($Corr\left(Y_{1,t-1}^{f}(1), Y_t\right) = Corr\left(Y_{2,t-1}^{f}(1), Y_t\right) = 0.152$), 2) A mid-correlation scenario ($Corr\left(Y_{1,t-1}^{f}(1), Y_t\right) = Corr\left(Y_{2,t-1}^{f}(1), Y_t\right) = 0.424$) and 3) A high-correlation scenario ($Corr\left(Y_{1,t-1}^{f}(1), Y_t\right) = Corr\left(Y_{2,t-1}^{f}(1), Y_t\right) = 0.770$).

For the low-correlation scenario, we set the following parameters: $\phi_y = 0.1, c = 0.2, \phi_x = 0.3, \sigma_\varepsilon^2 = 1, \sigma_u^2 = 2, Corr(\varepsilon_t, u_t) = 0.1, \beta_1 = 0.1, \beta_2 = 0.237, \sigma_\omega^2 = 1.5, \sigma_v^2 = 1, \mu_y = 0.15, \mu_x = 0.2, \alpha_1 = 0.1, \alpha_2 = 0.2, \phi_1 = 0.3, \phi_2 = 0.1, c_1 = 0.1, c_2 = 0.2, \mu_1 = 0.2$ and $\mu_2 = 0.1$. With these parameters both correlations $Corr\left(Y_{1,t-1}^{f}(1), Y_t\right)$ and $Corr\left(Y_{2,t-1}^{f}(1), Y_t\right)$ take the same value of 0.152 and therefore the null hypothesis is satisfied.

In the mid-correlation scenario, we set: : $\phi_y = 0.3, c = 0.5, \phi_x = 0.2, \sigma_\varepsilon^2 = 3, \sigma_u^2 = 2, Corr(\varepsilon_t, u_t) = 0.3, \beta_1 = 0.3, \beta_2 = 0.476, \sigma_\omega^2 = 3,\ \sigma_v^2 = 2, \mu_y = 0.1, \mu_x = 0.5, \alpha_1 = 0.3, \alpha_2 = 0.1, \phi_1 = 0.6, \phi_2 = 0.3, c_1 = 0.2, c_2 = 0.3, \mu_1 = 0.3$ and $\mu_2 = 0.1$. In this case, both correlations $Corr\left(Y_{1,t-1}^{f}(1), Y_t\right)$ and $Corr\left(Y_{2,t-1}^{f}(1), Y_t\right)$ take the same value of 0.424 and again the null hypothesis is satisfied.

Finally, in the high-correlation scenario, we set: : $\phi_y = 0.5, c = 0.7, \phi_x = 0.4, \sigma_\varepsilon^2 = 1, \sigma_u^2 = 1.5, Corr(\varepsilon_t, u_t) = 0.5, \beta_1 = 0.6, \beta_2 = 0.77, \sigma_\omega^2 = 2, \sigma_v^2 = 2.5, \mu_y = 0.3, \mu_x = 0.4, \alpha_1 = 0.1, \alpha_2 = 0.2, \phi_1 = 0.4, \phi_2 = 0.6, c_1 = 0.4, c_2 = 0.6, \mu_1 = 0.2$

and $\mu_2 = 0.4$. With these parameters, $Corr(Y^f_{1,t-1}(1), Y_t) = Corr(Y^f_{2,t-1}(1), Y_t) = 0.770$ and the null hypothesis is imposed again.

Table A1 next exhibit our results. Qualitatively speaking, Table A1 delivers similar conclusions than Table 1: i) both versions of our test are nicely sized for large samples, ii) the friendly user version tends to be more oversized than the complete test for small samples, and iii) our test performs similarly well for the three scenarios of correlations in the null hypothesis.

Table A1: Size analysis imposing the null hypothesis $H_0: Corr(Y^f_{1,t-1}(1), Y_t) = Corr(Y^f_{2,t-1}(1), Y_t)$ for three different scenarios of correlations.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Correlation scenario	Low Correlation		Mid Correlation		High Correlation	
Nominal Size	10%	5%	10%	5%	10%	5%
			Sample Size=2000			
Correlation test	9.58	4.66	9.40	4.58	9.76	4.98
Friendly-user test	9.88	4.84	9.44	4.66	10.04	5.32
			Sample Size=500			
Correlation test	10.58	5.28	10.54	4.94	9.94	4.96
Friendly-user test	11.04	5.66	10.68	5.12	10.56	5.46
			Sample Size=100			
Correlation test	12.02	5.66	12.16	6.74	11.16	5.42
Friendly-user test	13.32	7.04	12.74	7.58	13.20	7.64
			Sample Size=50			
Correlation test	12.00	5.78	14.26	7.88	10.20	4.58
Friendly-user test	14.86	8.46	15.8	9.26	14.64	8.16

Notes: Each entry represents the percentage of rejections of the null hypothesis of equal correlations. Columns (2), (4) and (6) consider a nominal size of 10%, while columns (3), (5) and (7) consider a nominal size of 5%. As both versions of our test are asymptotically normal, we reject the null hypothesis using standard normal critical values. We consider 10,000 Monte Carlo simulations for each exercise. The "friendly user" version of our test requires subtracting the mean of both forecasts and the target variable. Long-run variances are estimated using Newey and West (1987, 1994). Source: Author's elaboration.

Power analysis

Finally, we explore the power of our test in this parameterization using one-step-ahead forecasts. To this end, we consider as a start point the following parameters of our mid-correlations scenario from Table A1: $\phi_y = 0.3, c = 0.5, \phi_x = 0.2, \sigma^2_\varepsilon = 3, \sigma^2_u = 2, Corr(\varepsilon_t, u_t) = 0.3, \beta_1 = 0.3, \beta_2 = 0.476, \sigma^2_\omega = 3,$ $\sigma^2_v = 2, \mu_y = 0.1, \mu_x = 0.5, \alpha_1 = 0.3, \alpha_2 = 0.1, \phi_1 = 0.6, \phi_2 = 0.3, c_1 = 0.2, c_2 = 0.3, \mu_1 = 0.3$ and $\mu_2 = 0.1$. Recall that, under these parameters, we impose the null hypothesis at $Corr(Y^f_{1,t-1}(1), Y_t) = Corr(Y^f_{2,t-1}(1), Y_t) = 0.424$. To impose the alternative hypothesis, we shrink the coefficient β_2 to four different values: 0.1, -0.1, -0.25, and -0.4. With these new values of

β_2, the differences in correlations become $Corr\left(Y_{1,t-1}^{f}(1), Y_t\right) - Corr\left(Y_{2,t-1}^{f}(1), Y_t\right) = \rho_1 - \rho_2 = 0.059, 0.101, 0.136$ and 0.175 respectively. We consider four different sample sizes: 50, 100, 500, 1000, and 2000. Finally, we run 10,000 Monte Carlo replications with nominal sizes of 10% and 5%.

Table A2 exhibit our results. Qualitatively speaking, Table A2 delivers similar conclusions than Table 6: i) both versions of our test reject more than 90% in large samples (specially for T=2000), ii) obviously the power of our test increases with a greater difference between both correlations, iii) the power of our test tends to deteriorate with small samples, specially when the differences between both correlations are small, and iv) in small samples, the Friendly-user test has an edge in terms of power; this result should be seen with caution, since our simulations indicates that the Friendly-user test tends to be more oversized, specially in small samples.

Table A2: Power analysis imposing the alternative hypothesis for a mid-correlations scenario.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Under the alternative hypothesis:	$\rho_1 - \rho_2 = 0.059$		$\rho_1 - \rho_2 = 0.101$		$\rho_1 - \rho_2 = 0.136$		$\rho_1 - \rho_2 = 0.175$	
Nominal Size	10%	5%	10%	5%	10%	5%	10%	5%
				Sample Size = 2000				
Correlation test	99.84	99.52	100.00	100.00	100.00	100.00	100.00	100.00
Friendly-user test	99.84	99.52	100.00	100.00	100.00	100.00	100.00	100.00
				Sample Size = 1000				
Correlation test	94.67	90.07	99.98	99.94	100.00	100.00	100.00	100.00
Friendly-user test	94.63	89.92	99.98	99.94	100.00	100.00	100.00	100.00
				Sample Size = 500				
Correlation test	73.44	62.86	97.94	96.13	99.93	99.76	100.00	100.00
Friendly-user test	73.26	62.56	97.91	96.01	99.91	99.75	100.00	100.00
				Sample Size = 100				
Correlation test	28.31	19.00	51.12	38.90	68.74	57.09	81.01	71.61
Friendly-user test	28.48	19.23	51.02	38.60	68.52	56.64	80.77	71.13
				Sample Size = 50				
Correlation test	20.95	13.16	33.29	23.06	45.24	34.13	57.70	45.18
Friendly-user test	21.76	13.75	33.72	23.37	45.37	34.07	57.69	44.79

Notes: Each entry represents the percentage of rejections of the null hypothesis of equal correlations. The alternative hypothesis is set at four different levels: $\rho_1 - \rho_2 = 0.059, 0.101, 0.136$ and 0.175 respectively. Columns (2), (4), (6) and (8) consider a nominal size of 10%, while columns (3), (5), (7) and (9) consider a nominal size of 5%. As both versions of our test are asymptotically normal, we reject the null hypothesis using standard normal critical values. We consider 5,000 Monte Carlo simulations for each exercise. The "friendly user" version of our test requires subtracting the mean of both forecasts and the target variable. Long-run variances are estimated using Newey and West (1987, 1994). Source: Author's elaboration.

Appendix A4. Derivations using the Delta Method.

Our test.

Define the following sample moments vector:

$$\begin{bmatrix} m_Z \\ m_X \\ m_Y \\ m_{ZZ} \\ m_{XX} \\ m_{ZY} \\ m_{XY} \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} Z_t \\ X_t \\ Y_t \\ Z_t^2 \\ X_t^2 \\ Z_tY_t \\ X_tY_t \end{bmatrix}$$

Let $[\mu_Z, \mu_X, \mu_Y, \mu_{ZZ}, \mu_{XX}, \mu_{ZY}, \mu_{XY}]^T$ be the population counterpart. Additionally, let $s_y^2 = m_{YY} - m_Y^2, s_Z^2 = m_{ZZ} - m_Z^2, s_X^2 = m_{XX} - m_X^2, s_{YZ} = m_{YZ} - m_z m_y, s_{YX} = m_{YX} - m_X m_Y$.

By the Central Limit Theorem (CLT):

$$\sqrt{T}\left(\begin{bmatrix} m_Z \\ m_X \\ m_Y \\ m_{ZZ} \\ m_{XX} \\ m_{ZY} \\ m_{XY} \end{bmatrix} - \begin{bmatrix} \mu_Z \\ \mu_X \\ \mu_Y \\ \mu_{ZZ} \\ \mu_{XX} \\ \mu_{ZY} \\ \mu_{XY} \end{bmatrix}\right) \xrightarrow{d} N_7(0_{7x1}, \sum_{j=-\infty}^{\infty} \Omega_{\mathrm{j}})$$

Where Ω_{j} =

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

Now let us define the function $g: R^7 \to R^4$ such that

$$g\begin{bmatrix} m_Z \\ m_X \\ m_Y \\ m_{ZZ} \\ m_{XX} \\ m_{ZY} \\ m_{XY} \end{bmatrix} = \begin{bmatrix} s_Z^2 \\ s_X^2 \\ s_{YZ} \\ s_{YX} \end{bmatrix} = \begin{bmatrix} m_{ZZ} - m_Z^2 \\ m_{XX} - m_X^2 \\ m_{YZ} - m_Y m_Z \\ m_{YX} - m_Y m_X \end{bmatrix}$$

Let Δg be the first derivative of g, then:

$$\Delta g \begin{bmatrix} m_Z \\ m_X \\ m_Y \\ m_{ZZ} \\ m_{XX} \\ m_{ZY} \\ m_{XY} \end{bmatrix} = \begin{bmatrix} -2m_Z & 0 & -m_Y & 0 \\ 0 & -2m_X & 0 & -m_Y \\ 0 & 0 & -m_Z & -m_X \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Or applied to our population moments:

$$\Delta g \begin{bmatrix} \mu_Z \\ \mu_X \\ \mu_Y \\ \mu_{ZZ} \\ \mu_{XX} \\ \mu_{ZY} \\ \mu_{XY} \end{bmatrix} = \begin{bmatrix} -2\mu_Z & 0 & -\mu_Y & 0 \\ 0 & -2\mu_X & 0 & -\mu_Y \\ 0 & 0 & -\mu_Z & -\mu_X \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Hence by the Delta method

$$\sqrt{T}\left(\begin{bmatrix} s_Z^2 \\ s_X^2 \\ s_{YZ} \\ s_{YX} \end{bmatrix} - \begin{bmatrix} \sigma_Z^2 \\ \sigma_X^2 \\ \sigma_{YZ} \\ \sigma_{YX} \end{bmatrix}\right) \xrightarrow{d} N_4\left(0_{4x1}, \Delta g^T \left[\sum_{j=-\infty}^{\infty} \Omega_{\mathrm{j}}\right] \Delta g\right)$$

Now let $h: R^4 \to R$ such that

$$h\begin{bmatrix} s_Z^2 \\ s_X^2 \\ s_{YZ} \\ s_{YX} \end{bmatrix} = \frac{s_{YZ}}{s_Z} - \frac{S_{YX}}{s_X}$$

Hence the first derivative of h evaluated at our population moments:

$$\Delta h \begin{bmatrix} \sigma_Z^2 \\ \sigma_X^2 \\ \sigma_{YZ} \\ \sigma_{YX} \end{bmatrix} = \begin{bmatrix} -\frac{\sigma_{YZ}}{2\sigma_Z^3} \\ \frac{\sigma_{YX}}{2\sigma_X^3} \\ \frac{1}{\sigma_Z} \\ -\frac{1}{\sigma_X} \end{bmatrix}$$

And the desired result follows from using the Delta method once more:

$$\sqrt{T}s_Y([r_1 - r_2] - [\rho_1 - \rho_2]) \xrightarrow{d} N\left(0, \Delta h^T \Delta g^T \left[\sum_{j=-\infty}^{\infty} \Omega_j\right] \Delta g \Delta h\right) \blacksquare$$

Our friendly-user test.

Define the following sample moments vector:

$$\begin{bmatrix} m_Z \\ m_X \\ m_Y \\ m_{ZZ} \\ m_{XX} \\ m_{ZY} \\ m_{XY} \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} Z_t \\ X_t \\ Y_t \\ Z_t^2 \\ X_t^2 \\ Z_t Y_t \\ X_t Y_t \end{bmatrix}$$

Let $[\mu_Z, \mu_X, \mu_Y, \mu_{ZZ}, \mu_{XX}, \mu_{ZY}, \mu_{XY}]^T$ be the population counterpart. Additionally, let $s_y^2 = m_{YY} - m_Y^2, s_Z^2 = m_{ZZ} - m_Z^2, s_X^2 = m_{XX} - m_X^2, s_{YZ} = m_{YZ} - m_z m_y, s_{YX} = m_{YX} - m_X m_Y$.

By the Central Limit Theorem (CLT):

$$\sqrt{T}\left(\begin{bmatrix} m_Z \\ m_X \\ m_Y \\ m_{ZZ} \\ m_{XX} \\ m_{ZY} \\ m_{XY} \end{bmatrix} - \begin{bmatrix} \mu_Z \\ \mu_X \\ \mu_Y \\ \mu_{ZZ} \\ \mu_{XX} \\ \mu_{ZY} \\ \mu_{XY} \end{bmatrix}\right) \xrightarrow{d} N_7(0_{7x1}, \sum_{j=-\infty}^{\infty} \Omega_j)$$

Where Ω_j =

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

Now let us define the function $g: R^7 \to R^4$ such that

$$g\begin{bmatrix} m_Z \\ m_X \\ m_Y \\ m_{ZZ} \\ m_{XX} \\ m_{ZY} \\ m_{XY} \end{bmatrix} = \begin{bmatrix} s_Z^2 \\ s_X^2 \\ s_{YZ} \\ s_{YX} \end{bmatrix} = \begin{bmatrix} m_{ZZ} - m_Z^2 \\ m_{XX} - m_X^2 \\ m_{YZ} - m_Y m_Z \\ m_{YX} - m_Y m_X \end{bmatrix}$$

Let Δg be the first derivative of g, then:

$$\Delta g\begin{bmatrix} m_Z \\ m_X \\ m_Y \\ m_{ZZ} \\ m_{XX} \\ m_{ZY} \\ m_{XY} \end{bmatrix} = \begin{bmatrix} -2m_Z & 0 & -m_Y & 0 \\ 0 & -2m_X & 0 & -m_Y \\ 0 & 0 & -m_Z & -m_X \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Or applied to our population moments:

$$\Delta g\begin{bmatrix} \mu_Z \\ \mu_X \\ \mu_Y \\ \mu_{ZZ} \\ \mu_{XX} \\ \mu_{ZY} \\ \mu_{XY} \end{bmatrix} = \begin{bmatrix} -2\mu_Z & 0 & -\mu_Y & 0 \\ 0 & -2\mu_X & 0 & -\mu_Y \\ 0 & 0 & -\mu_Z & -\mu_X \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

But in this case, we have assumed that $\mu_Z = \mu_X = \mu_Y = 0$, hence

$$\Delta g\begin{bmatrix} 0 \\ 0 \\ 0 \\ \mu_{ZZ} \\ \mu_{XX} \\ \mu_{ZY} \\ \mu_{XY} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

And by the Delta method

$$\sqrt{T}\left(\begin{bmatrix} s_Z^2 \\ s_X^2 \\ s_{YZ} \\ s_{YX} \end{bmatrix} - \begin{bmatrix} \sigma_Z^2 \\ \sigma_X^2 \\ \sigma_{YZ} \\ \sigma_{YX} \end{bmatrix}\right) \xrightarrow{d} N_4(0_{4x1}, \sum_{j=-\infty}^{\infty} \Gamma_j)$$

$$\sum_{j=-\infty}^{\infty} \Gamma_{\mathrm{j}} = \Delta g^{T}\left[\sum_{j=-\infty}^{\infty} \Omega_{\mathrm{j}}\right]\Delta g =$$

$$\sum_{j=-\infty}^{\infty}\begin{bmatrix} Cov(Z_t^2, Z_{t-j}^2) & Cov(X_t^2, Z_{t-j}^2) & Cov(Z_tY_t, Z_{t-j}^2) & Cov(X_tY_t, Z_{t-j}^2) \\ Cov(Z_t^2, X_{t-j}^2) & Cov(X_t^2, X_{t-j}^2) & Cov(Z_tY_t, X_{t-j}^2) & Cov(X_tY_t, X_{t-j}^2) \\ Cov(Z_t^2, Z_{t-j}Y_{t-j}) & Cov(X_t^2, Z_{t-j}Y_{t-j}) & Cov(Z_tY_t, Z_{t-j}Y_{t-j}) & Cov(X_tY_t, Z_{t-j}Y_{t-j}) \\ Cov(Z_t^2, X_{t-j}Y_{t-j}) & Cov(X_t^2, X_{t-j}Y_{t-j}) & Cov(Z_tY_t, X_{t-j}Y_{t-j}) & Cov(X_tY_t, X_{t-j}Y_{t-j}) \end{bmatrix}$$

Now let $h: R^4 \rightarrow R$ such that

$$h\begin{bmatrix} s_Z^2 \\ s_X^2 \\ s_{YZ} \\ s_{YX} \end{bmatrix} = \frac{s_{YZ}}{s_Z} - \frac{S_{YX}}{s_X}$$

Hence the first derivative of h evaluated at our population moments:

$$\Delta h\begin{bmatrix} \sigma_Z^2 \\ \sigma_X^2 \\ \sigma_{YZ} \\ \sigma_{YX} \end{bmatrix} = \begin{bmatrix} -\dfrac{\sigma_{YZ}}{2\sigma_Z^3} \\ \dfrac{\sigma_{YX}}{2\sigma_X^3} \\ \dfrac{1}{\sigma_Z} \\ -\dfrac{1}{\sigma_X} \end{bmatrix}$$

And the desired result follows from using the Delta method once more:

$$\sqrt{T}s_Y([r_1 - r_2] - [\rho_1 - \rho_2]) \xrightarrow{d} N\left(0, \nabla\hat{h}'[\sum_{j=-\infty}^{\infty} \hat{\Gamma}_{\mathrm{j}}]\,\nabla\hat{h}\right) \blacksquare$$