Economic Growth Centre
Economics, School of Social Sciences
Nanyang Technological University
14 Nanyang Drive
Singapore 637332

COULD THE CORRELATION OF A STATIONARY SERIES WITH A NON-STATIONARY SERIES OBTAIN MEANINGFUL OUTCOMES?

Wing-Keung Wong
Minh Tam Pham

October 2025

EGC Report No: 2025/09

HSS-04-90A
Tel: +65 67906073
Email: D-EGC@ntu.edu.sg

Economic Growth Centre Working Paper Series

Could the correlation of a stationary series with a non-stationary series obtain meaningful outcomes? *

Wing-Keung Wong

Department of Finance, Quantum AI Research Center, Fintech & Blockchain Research Center, and Big Data Research Center, Asia University

Department of Medical Research, China Medical University Hospital

Business, Economic and Public Policy Research Centre, Hong Kong Shue Yan University

The Economic Growth Centre, Nanyang Technological University

Minh Tam Pham

Department of Finance, College of Management, Asia University

October 25, 2025

Abstract:

While many studies report correlations between a stationary time series Y_t

*The first author would like to thank Robert B. Miller and Howard E. Thompson for their continuous guidance and encouragement. The first author would also like to thank George Edward Pelham Box and Gregory C. Reinsel for teaching him time series analysis. We thank Professor Mu Yue for giving us her program for Wong and Yue (2024) so that we can develop the program for this paper. This paper was done when the first author visited the Business, Economic and Public Policy Research Centre and The Economic Growth Centre, Nanyang Technological University in 2025. This research has been supported by Asia University, China Medical University Hospital, Hong Kong Shue Yan University, Nanyang Technological University, Research Grants Council (RGC) of Hong Kong (project numbers 12502814 and 12500915), and the Ministry of Science and Technology (MOST, Project Numbers 106-2410-H-468-002 and 107-2410-H-468-002-MY3), Taiwan.

and a non-stationary time series X_t, it is still an open problem whether traditional correlation tests are appropriate for assessing the significance of such relationships. To address this gap, we first hypothesize that applying standard regression-based correlation tests in this context may yield spurious or non-informative results. Furthermore, we conjecture that standard correlation statistics are not suitable for evaluating such relationships, and the appropriate test statistic differs from that used for stationary or jointly random series.

We first validate our conjectures through simulation studies. In our experiments, Y_t follows a stationary AR(1) process with parameter ϕ, while X_t follows a random-walk model. The empirical rejection rate exceeds the nominal 5% level, increasing from 7.86% when $\phi = 0.1$ to around 62.3% when $\phi = 0.9$. Our findings support our claims about the spurious nature of the correlation and the inadequacy of standard tests in this setting.

Thereafter, we develop the estimation and testing theory for the correlation between a stationary Y_t and a non-stationary X_t. We have proved that the standard correlation statistic cannot be used in this setting and that the resulting test statistic differs from the one used to test the correlation between two random series Y_t and X_t, concluding that the traditional correlation test cannot be used to test for the correlation between a stationary time series Y_t and a non-stationary time series X_t.

Keywords: Cointegration, stationarity, non-stationarity, correlation , time series analysis

JEL Classification: C01, C15, C22, C58, C60

1 Introduction

We read many papers in the literature and found that some papers report results of the correlation of a stationary Y_t with a non-stationary X_t. For example, Singh et al. (2011) find that GDP (which is not stationary) has positive relationships with stock returns (which is stationary). Other papers study the correlation of GDP (which is not stationary) with some variables that could be stationary; see, for example, Lakstutiene (2008), Vinkler (2008), Andrei, *et al.* (2009), Valadez (2011), Szigeti, *et al.* (2013), Marcu, *et al.* (2015), Bilyuga, *et al.* (2016), Anghelache, *et al.* (2019), and many others. However, as far as we know, there is no paper telling readers whether we can use the traditional correlation test to test whether the correlation of a stationary time series with a non-stationary time series is significant.

Thus, to bridge the gap in the literature, in this paper, we first conjecture that the correlation of a stationary Y_t with a non-stationary X_t could be spurious if one uses the tests from the standard regression model, and the correlation of a stationary Y_t with a non-stationary X_t may not be able to get any meaningful outcome. For comparison, we also conjecture that traditional correlation tests can be applied to the correlations between two random series Y_t and X_t. At last, we set the conjecture to see whether the distributions of the test statistics are different for the two cases. We then conjecture that the tests from the standard correlation statistic cannot be used to test a correlation of a stationary Y_t with a non-stationary X_t, and the statistic obtained for testing a correlation of a stationary Y_t with a non-stationary X_t is different from that for testing the correlation of two random series Y_t and X_t.

To examine whether all of our conjectures hold, we first conduct simulations to examine whether the conjectures hold. We then develop some theorems to confirm whether the conjectures hold. To examine whether the first three conjectures hold, we set up a stationary AR(1) model for Y_t and a non-stationary model for X_t. We then develop the algorithm for the simulations of getting the correlation of a stationary Y_t with a non-stationary X_t and conduct simulations to obtain 10,000 pairs of X_t and Y_t for each pair of the AR parameter ϕ and the sample size N. For each pair of simulated X_t and Y_t, we obtain the traditional correlation coefficient, $\hat{\rho}_{xy}$, between X and Y and use the T-test to obtain the statistic and its corresponding p-value. We then compare each p-value with 0.05 and compute the proportion of p-values smaller than 0.05 among the 10,000 p-values. We denote this as the *rejection rate* in the table. For comparison, we set up similar conjectures for the case with two random series Y_t and X_t, develop the algorithm, and conduct simulations getting the correlation of two random series Y_t and X_t.

Our simulations show that when Y_t follows a stationary AR(1) model with the AR parameter ϕ and X_t follows a random-walk model, the rejection rate is more than 5% and increases as ϕ increases (from around 7.86% when $\phi = 0.1$ to around 62.3% when $\phi = 0.9$), on average. This implies that the first three conjectures hold and leads us to conclude that when Y_t follows a stationary AR(1) model and X_t follows a random-walk model, then the correlation of Y_t and X_t may not provide any meaningful outcome, the correlation could be spurious, and the standard correlation test may not be suitable for testing a correlation between Y_t and X_t. On the other hand, our simulations also find that when Y_t is a random series and X_t follows a random-walk model, the rejection rate is close to 5%, implying that the first three conjectures could still hold. In this case, the correlation of Y_t and X_t may provide a meaningful outcome, may not be spurious, and the standard correlation test could be used to test whether the correlation of Y_t and X_t is zero. Moreover, we also find that when both Y_t and X_t follow random-walk models, the rejection rate is 86.29%

on average, which is much more than 5%. This implies that all first three conjectures hold in this case, and leads us to conclude that when both Y_t and X_t follow random-walk models, the correlation of Y_t and X_t may not provide any meaningful outcome, the correlation could be spurious, and the test from the standard correlation may not be suitable.

Thereafter, we develop the estimation and testing theory for getting the correlation of a stationary Y_t with a non-stationary X_t and prove that the test from the standard correlation statistic cannot be used to test a correlation of a stationary Y_t with a non-stationary X_t and the statistic obtained for testing a correlation of a stationary Y_t with a non-stationary X_t is different from that for testing the correlation of two random series Y_t and X_t.

Section 2 provides background literature on the topic. Section 3 discusses the theory of the standard correlation coefficient and traditional correlation test. Section 4 discusses the model setup for the simulation, develops an algorithm for the simulations, and discusses the simulation results by using the algorithm developed in our paper. Section 5 develops the estimation and testing theory for getting the correlation of a stationary series with a non-stationary series and proves that the test from the standard correlation statistic cannot be used to test a correlation of a stationary series with a non-stationary series and the statistic obtained for testing a correlation of a stationary series with a non-stationary series is different from that for testing the correlation of two random series. The last section concludes and suggests future extensions.

2 Literature Review

Time series analysis involves understanding and modeling sequences of data points collected over time at regular intervals. Over the years, various models have been proposed for modeling time series data. Readers may refer to (Tsay, 1989; Nakatani and Teräsvirta, 2009) and others for more information. These models assume that the underlying time series is either stationary or becomes stationary after differencing or other transformations (Brockwell and Davis, 2002). However, when time series data are non-stationary, models such as ARIMA are commonly used, as non-stationary series can present challenges in statistical inference.

One of the key challenges in time series econometrics is the issue of spurious regression, as first noted by Granger and Newbold (1974), who demonstrated that regressing two independent non-stationary time series often results in highly significant coefficients and high R^2 values, despite no actual causal relationship. This issue was later extended by Phillips (1986), Sun (2004), and others, who provided asymptotic results for spurious regressions. Subsequent works by Ventosa-Santaulária (2009), Marmol (1995), and Kao (1999) further developed the theory and diagnostic tools for detecting spurious regressions, with particular focus on $I(1)$ and fractionally integrated ($I(d)$) processes with d satisfying $0 < d < 1$.

Many studies have highlighted the danger of regressing independent non-stationary series, as it often leads to misleading conclusions (Granger and Newbold, 1974; Phillips, 1986). Agiakloglou (2013) examined spurious regressions in the context of both stationary and non-stationary series, while Kim et al. (2004) explored spurious outcomes when one series has a linear trend and the other is stationary. The traditional remedy to avoid spurious regression is either differencing for the non-stationary series or using cointegration techniques when the series are related but exhibit non-stationarity (Engle and Granger, 1987).

Several studies focus on spurious regression with mixed integration orders. For instance, Pesaran et al. (1999) and Westerlund (2008) studied scenarios where a combination of stationary ($I(0)$) and non-stationary ($I(1)$) series may lead to spurious results. The issue of spurious regression has also been extended to long-memory processes by Tsay and Chung (2000), who considered fractionally integrated ($I(d)$) series, and to scenarios involving seasonal unit roots, as explored by Abeysinghe (1994).

Recently, Cheng et al. (2021) conducted simulations and found that, in some situations, the regression of two independent and nearly non-stationary series does not exhibit spurious problems at all. Cheng et al. (2022) found cases where regression is insignificant even though the variables are actually related. On the other hand, Wong et al. (2024) showed through simulations that regression of stationary time series can still be spurious and proposed a remedial approach to address the issue.

In the literature, some papers report results of regressing a stationary time series on a non-stationary time series but no paper examines the robustness of inference in such settings. To bridge the gap in the literature, Wong and Yue (2024) conducted a simulation and found that regressing a stationary time series on a non-stationary time series could be spurious. Thereafter, they developed the estimation and testing theory and concluded that the traditional statistic T_N^β for testing $H_0^\beta : \beta = \beta_0$ versus $H_1^\beta : \beta \neq \beta_0$ does not have any asymptote distribution with $E(T_N^\beta) \to \infty$ and $Var(T_N^\beta) \to \infty$ as $N \to \infty$, and thus, it cannot be used for testing the regression of a stationary time series on a non-stationary time series.

3 The Models

In this paper, we investigate whether getting the correlation of a stationary time series, Y_t, with a non-stationary time series, X_t, could obtain meaningful outcomes. To do so, in this section, we first discuss the standard correlation coefficient and traditional correlation test in the following subsection.

3.1 Correlation

For any two variables X and Y, the **correlation coefficient** between two variables X and Y is defined as:

$$\rho_{xy} = \frac{Cov(X,Y)}{\sqrt{Var\,(X)\,Var\,(Y)}}\ , \tag{3.1}$$

in which $Cov\,(X,Y)$ is the **covariance** of X and Y such that

$$Cov\,(X,Y) = E\,[X - E\,(X))\,(Y - E\,(Y)] = E\,(XY) - E\,(X)\,E\,(Y)\ .$$

To estimate ρ_{xy}, one can get N pairs of data (x_i, y_i) $(I = 1, \cdots, N)$ for (X,Y) and use the following traditional correlation coefficient, r_{xy}, between X and Y:

$$\begin{aligned} r_{xy} &= \hat{\rho}_{xy} = \frac{s_{xy}}{s_x s_y} \\ &= \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2 \sum(Y_i - \overline{Y})^2}}\ , \end{aligned} \tag{3.2}$$

where Cov$(X,Y) = \sigma_{XY}$ is estimated by the sample covariance, s_{xy}, such that

$$s_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y}), \tag{3.3}$$

Var(X) is the variance of X which can be estimated by the sample variance of X, s_x^2, such that

$$s_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2\ , \tag{3.4}$$

and Var(Y) is the variance of Y which can be estimated by the sample variance of Y, s_y^2, such that

$$s_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y})^2, \tag{3.5}$$

To test whether there is no correlation between X and Y, we use

$$H_0 \; : \; \rho = 0 \qquad \text{versus} \qquad H_1 \; : \; \rho \neq 0 \; . \tag{3.6}$$

If the null hypothesis H_0 in Equation (3.6) is true, it is well-known that one can use the following proposition:

Proposition 3.1 *In the above model setting, if the null hypothesis H_0 in Equation* (3.6) *is true, we have*

$$T_N = \frac{r_{xy}}{s_r} \xrightarrow{d} T_{(N-2)} \; , \tag{3.7}$$

with $s_r = \sqrt{\frac{1-r_{xy}^2}{N-2}}$ and $T_{(N-2)}$ is the t-distribution with $N-2$ degrees of freedom.

One can use T_N to test whether the null hypothesis H_0 in Equation (3.6) holds.

3.2 Correlation of a stationary Y_t with a non-stationary X_t

We read many papers in the literature and found that some papers report results of the correlation of a stationary Y_t with a non-stationary X_t. For example, Singh et al. (2011) find that GDP (which is not stationary) has positive relationships with stock returns (which are stationary). Some papers study the correlation of GDP (which is not stationary) with some variables that could be stationary, see, for example, Lakstutiene (2008), Vinkler (2008), Andrei, *et al.* (2009), Valadez (2011), Szigeti, *et al.* (2013), Marcu, *et al.* (2015), Bilyuga, *et al.* (2016), Anghelache, *et al.* (2019), and many others. Thus, in this paper, we investigate whether there is any problem in getting the correlation of a stationary time series with a non-stationary time series by examining the following conjecture:

Conjecture 3.1 *Correlation of a stationary Y_t with a non-stationary X_t may not yield any meaningful outcome.*

Before we examine the above conjecture, we first examine the following conjecture:

Conjecture 3.2 *Correlation of a stationary Y_t with a non-stationary X_t could be spurious if one uses the tests from the standard regression model as shown in Section 3.1.*

In addition, we set the following conjecture:

Conjecture 3.3 *The tests from the standard correlation, as shown in Section 3.1, cannot be used to test a correlation of a stationary Y_t with a non-stationary X_t.*

3.3 Correlation of two random series

In order to compare the results obtained in Section 3.2 in which the traditional correlation test may not be able to be used, in this section, we will study the correlation of two random series Y_t and X_t so that we can see whether the distribution of the test statistics are different for these two cases. To do so, we first set the following conjectures:

Conjecture 3.4 *Correlation of two random series Y_t and X_t will lead to get meaningful outcome.*

Conjecture 3.5 *Correlation of two random series Y_t and X_t is not spurious if one uses the tests from the standard regression model as shown in Section 3.1.*

Conjecture 3.6 *The tests from the standard correlation, as shown in Section 3.1, can be used to test a correlation of two random series Y_t and X_t.*

At least, we set the following conjecture to see whether the distributions of the test statistics are different for the two cases:

Conjecture 3.7 *The statistic obtained for testing a correlation of a stationary Y_t with a non-stationary X_t is different from that for testing the correlation of two random series Y_t and X_t.*

To examine whether the correlation of a stationary time series to a non-stationary time series could yield any meaningful result, we first conduct a simulation to examine whether all the conjectures hold in the next section, and thereafter, study whether we prove that the conjectures hold in Section 5.

4 Model Setup and Simulation

To study whether Conjectures 3.1 to 3.7 hold, we first set up a model for the simulation as shown in this section. We then develop an algorithm for our simulations and discuss the simulation results by using the algorithm developed in our paper.

4.1 Model Setup

We first discuss the model set-up for the simulation in the following subsection:

4.1.1 Correlation of a stationary Y_t with a non-stationary X_t

In this paper, we consider a stationary series to be weakly stationary or covariance stationary. One could easily extend our result to include strictly stationary series. We will use the following linear first-order autoregressive AR(1):

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2) \ . \tag{4.1}$$

To examine whether Conjectures 3.1 and 3.2 hold, we will compute the correlation of the stationary Y_t defined in Equation (4.1) with a non-stationary X_t such that

$$X_t = X_{t-1} + e_t, \quad e_t \overset{iid}{\sim} N(0, \sigma_e^2) \ . \tag{4.2}$$

We note that one could use more complicated distributions for both ε_t and e_t. In this paper, we only consider a simple case in which both ε_t and e_t are normally distributed and ε_i and e_j are independent for any i and j. We now impose the following conditions to X_t and Y_t:

$$X_0 = 0 \quad \text{and} \quad Y_0 = 0 \; . \tag{4.3}$$

We will use Equation (4.1) to define Y_t and use Equation (4.2) to define X_t. We note that one could use more complicated distributions for both ε_t and e_t. In this paper, we only consider a simple case in which both ε_t and $e_t \sim$ N(0,1). However, in order to simulate X_t and Y_t properly, without loss of generality, we will consider different factors that may affect the properties of time series carefully.

First, we control the lengths of time series. Because longer series will include more information than shorter ones, in this paper, we simulate the time series with the following four different lengths in our study:

(i) N=100, (ii) N=500, (iii) N=1000.

After deciding the lengths of both X_t and Y_t, we now consider the different values of ϕ as follows:

(a) $\phi = 0$, (b) $\phi = 0.1$, (c) $\phi = 0.3$, (d) $\phi = 0.5$, (e) $\phi = 0.7$, (f) $\phi = 0.9$.

For simplicity, we only consider the positive coefficient of ϕ in this paper.

With 3 different time series lengths, and the above 6 combinations of ϕ values, there are in total 18 subcases of simulation in our study.

4.1.2 Correlation of two random series

In this section, we will use a random X_t such that

$$Y_t = \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2), \tag{4.4}$$

and use a random X_t such that

$$X_t = e_t, \quad e_t \overset{iid}{\sim} N(0, \sigma_e^2) \ . \tag{4.5}$$

This model setting is used to examine whether Conjecture 3.7 holds.

4.2 Algorithm

We turn to develop an algorithm for our simulations.

We first discuss the algorithm of the simulations for getting the correlation of a stationary Y_t with a non-stationary X_t as shown in the next subsection.

4.2.1 Correlation of a stationary Y_t with a non-stationary X_t

Since X_t and Y_t are generated independently, they are not related, and thus, H_0 in Equation (3.6) is expected not to be rejected. To test whether this is true, we will conduct simulations for each subcase (for different combinations of ϕ and N) described in Section 4.1. To do so, we develop the following algorithm for this purpose:

1. We first simulate 10,000 pairs of X_t and Y_t defined in Equations (4.2) and (4.1), respectively, for each pair ϕ and N described in Section 4.1.

2. For each pair of simulated X_t and Y_t, we use Equation (3.2) to obtain the traditional correlation coefficient, $\hat{\rho}_{xy}$, between X and Y and use the T_N test defined in Equation (3.7) to obtain the statistic and its corresponding p-value. Thus, for 10,000 pairs of X_t and Y_t in each subcase, we will obtain 10,000 $\hat{\rho}_{xy}$s and 10,000 corresponding p-values.

3. For each pair of simulated X_t and Y_t, we will compare their p-value with 0.05. If the p-value for the T_N test is more than 0.05, H_0 in Equation (3.6) is not rejected. In this sense, we compute the proportion of p-values that

are smaller than 0.05 among the 10,000 p-values and denote it as *rejection rate* in Table 4.1.

The above algorithm can be used to help to examine whether the T_N statistic in Equation (3.7) for the traditional correlation coefficient stated in Equation (3.2) follows a Student t-distribution. If X_t and Y_t are unrelated, H_0 in Equation (3.6) should be rejected around 5% of all the 10,000 simulations when the significance level is 0.05. Thus, if $\hat{\rho}_{xy}$s follows student t-distribution with $N-2$ degrees of freedom and the test is perfect, then the rejection rate should be exactly 0.05. If the rejection rate is significantly different from 0.05, then we can conclude that the t-test cannot be used and/or the correlation is spurious.

4.2.2 Correlation of two random series

We turn to discuss the algorithm of the simulations of getting the correlation of two random series Y_t and X_t. The algorithm is the same as the algorithm of getting the correlation of a stationary Y_t with a non-stationary X_t except in this case X_t and Y_t defined in Equations (4.5) and (4.4), respectively. Thus, we skip the discussion.

4.3 Simulation Results

We now discuss the simulation results by using the algorithms developed in our paper and stated in Section 4.2.1.

4.3.1 Correlation of a stationary Y_t with a non-stationary X_t

We first conduct simulation, obtain results by using the algorithm developed in our paper and stated in Section 4.2.1 for getting the correlation of a stationary Y_t with a non-stationary X_t and present the rejection rates in Table 4.1.

From Table 4.1, we first obtain the following observation:

Table 4.1: Rejection Rate in the simulation for various ϕ and N.

Case	Coefficients	N=100	N=500	N=1000	Average
(a)	$\phi = 0$	0.0488	0.0496	0.0514	0.0499
(b)	$\phi = 0.1$	0.0753	0.0771	0.0735	0.0753
(c)	$\phi = 0.3$	0.1368	0.1443	0.1491	0.1434
(d)	$\phi = 0.5$	0.2447	0.2565	0.2660	0.2557
(e)	$\phi = 0.7$	0.3743	0.4105	0.4055	0.3968
(f)	$\phi = 0.9$	0.5897	0.6344	0.6410	0.6217
(g)	$\phi = 1$	0.7685	0.8923	0.9278	0.8629

Note: The table summarizes the average rejection rates for each combination of sample size (n) and autoregressive coefficient (ϕ), calculated over 10,000 simulations, ϕ is the autoregressive coefficient for Y_t, and the rejection Rate is the average proportion of simulations in which the null hypothesis ($H_0 : \rho = 0$) was rejected, based on a p-value threshold of 0.05.

Observation 1 *When Y_t follows a stationary AR(1) model with the AR parameter ϕ, as defined in Equation (4.1), while X_t follows a random-walk model as defined in Equation (4.2), the rejection rate is more than 5% and increases as ϕ increases from around 7.86% when $\phi = 0.1$ to around 62.3% when $\phi = 0.9$) on average. This suggests that all Conjectures 3.1, 3.2, and 3.3 hold and conclude the following when Y_t follows a stationary AR(1) model and X_t follows a random-walk model:*

- *(1A) the correlation between Y_t and X_t may not yield any meaningful results,*
- *(1B) the correlation could be spurious, and*
- *(1C) the standard correlation tests, as described in Section 3.1, may not be appropriate for testing the correlation between Y_t and X_t.*

From Table 4.1, we also obtain the following observation:

Observation 2 *When Y_t is a random series and X_t follows a random-walk model, that is Case (a) in Table 4.1 when $\phi = 0$), the rejection rate is close to 5%. This indicates that Conjectures 3.1, 3.2, and 3.3 may still hold and we can conclude the following when Y_t is a random series and X_t follows a random-walk model,*

- *(2A) the correlation between Y_t and X_t yields meaningful results,*
- *(2B) the correlation is not spurious, and*
- *(2C) the standard correlation tests, as outlined in Section 3.1, can be used to test the correlation between Y_t and X_t.*

Last, from Table 4.1, we obtain the following observation:

Observation 3 *When both Y_t and X_t follow random-walk models defined in* (4.2)*, that is Case (f) in Table 4.1 when $\phi = 1$, the rejection rate is 86.29% on average, far exceeding 5%. This suggests that all Conjectures 3.1, 3.2, and 3.3 hold in this scenario and we can conclude the following when both Y_t and X_t follow random-walk models defined in* (4.2)*:*

- *(3A) the correlation between Y_t and X_t may not yield any meaningful results,*
- *(C2) the correlation could be spurious. and*
- *(C3) the standard correlation tests, as described in Section 3.1, may not be appropriate for testing the correlation between Y_t and X_t.*

In this paper, we will analyze Observation 1 and leave the analysis of Observations 2 and 3 in the future study.

4.3.2 Correlation of two random series

We then conduct simulation, obtain results by using the algorithm developed in our paper and stated in Section 4.2.2 for getting the correlation of two random series Y_t and X_t and present the rejection rates in Table 4.2.

Table 4.2: Rejection Rate in the simulation for various N.

$N = 100$	$N = 500$	$N = 1000$	Average
0.0529	0.0526	0.0490	0.0515

Note: The table presents the average rejection rate for each sample size (n), calculated over 10,000 simulations. In each simulation, two independent random variables (X_t and Y_t), both generated from normal distributions with mean 0 and standard deviation 1, are tested for correlation. The rejection rate represents the proportion of simulations in which the null hypothesis ($H_0 : \rho = 0$) was rejected based on a p-value threshold of 0.05.

The rejection rates in Table 4.2 are all close to the nominal level of 5%, concluding that Conjectures 3.4, 3.5, and 3.6 hold such that correlation of two random series Y_t and X_t will to get a meaningful outcome, correlation of two random series Y_t and X_t is not spurious if one uses the tests from the standard regression model as shown in Section 3.1, and the tests from the standard correlation as shown in Section 3.1 can be used to test a correlation of two random series Y_t and X_t. The results indicate that the correlation test is well-calibrated and performs as expected when testing for a correlation between two independent random series, X_t and Y_t. The small variations in rejection rates across different sample sizes suggest that the test's power does not dramatically change as the sample size increases, which is typical for tests involving independent random variables where the null hypothesis ($H_0 : \rho = 0$) is true. Overall, the results support the validity of the test in accurately identifying the absence of correlation, without producing false positives, even as the sample size varies.

The results of the simulation lead us to conclude that Conjectures 3.4, 3.5, and 3.6 hold such that the correlation of two random series Y_t and X_t will to get meaningful outcome, the correlation of two random series Y_t and X_t is not spurious if one uses the tests from the standard regression model as shown in Section 3.1, and the tests from the standard correlation as shown in Section 3.1 can be used to test a correlation of two random series Y_t and X_t.

5 The Theory

In this paper, our main objective is to develop the estimation and testing theory for getting the correlation of a stationary Y_t with a non-stationary X_t defined by Equations (4.1) and (4.2), respectively. Before we develop the theory, we first discuss the estimation and testing theory for getting the correlation of two random variables Y_t and X_t defined by Equations (4.4) and (4.5), respectively. To distinguish T_N defined in Equation (3.7) when the random variables Y_t and X_t defined by Equations (4.4) and (4.5), respectively, we make the following definition:

Definition 5.1 *We denote T_N be $\mathbf{T}_N$ defined in Equation (3.7) when Y_t and X_t are defined by Equations (4.1) and (4.2), respectively.*

The estimation and testing theory for getting the correlation of two random variables Y_t and X_t has been well-established, see Proposition 3.1. From Proposition 3.1, one can tell that $E(T_N) = 0$ and $Var(T_N) = 1$. Thus, in this paper, we will establish the results for $E(T_N) = 0$ and $Var(T_N) = 1$ in the next subsection so that one can use the results to compare the results of the estimation and testing theory for getting the correlation of a stationary Y_t with a non-stationary X_t.

5.1 Correlation of two random series

In this section, we first develop the estimation and testing theory for getting the correlation of two random variables Y_t and X_t defined by Equations (4.1) and (4.2), respectively. We first make the following assumption in this paper:

Assumption 5.1 *Under the model setting stated in Section 4.1, in this paper, we assume that e_i and ε_j are independent for any i and j.*

We now develop the following theorem:

Theorem 5.1 *Under the model setting stated in Section 4.1, if Y_t and X_t are defined by Equations* (4.4) *and* (4.5)*, respectively, and Assumption 5.1 holds, then $E(T_N) = 0$ where T_N is defined in* (3.7)*.*

Proof. Refer to Wong and Pham (2025). ■

We note that the approximation in Equation (**??**) holds because $\bar{e}_t$ and $\sum_{t=1}^{N} e_t^2$ are independent and $\bar{\varepsilon}_t$ and $\sum_{t=1}^{N} \varepsilon_t^2$ are independent, see, for example, Greene (2012), and Assumption 5.1 holds.

We turn to develop the following theorem for $Var(T_N)$ for 2 random variables:

Theorem 5.2 *Under the model setting stated in Section 4.1, if Y_t and X_t are defined by Equations* (4.4) *and* (4.5)*, respectively, and Assumption 5.1 holds, then $Var(T_N) = 1$ where T_N is defined in* (3.7)*.*

Proof. Refer to Wong and Pham (2025). ■

5.2 Correlation of a stationary Y_t with a non-stationary X_t

We turn to develop the estimation and testing theory for getting the correlation of a stationary Y_t with a non-stationary X_t in this section. Before we develop the theory, we first state the following proposition:

Proposition 5.1 *Under the model setting stated in Section 4.1, without loss of generality, we assume* $X_0 = 0$ *and* $Y_0 = 0$*, then we have*

$$X_t = \sum_{i=1}^{t} e_i \quad and \quad Y_t = \sum_{j=1}^{t} \phi^{t-j} \varepsilon_j \ . \tag{5.1}$$

We make the following assumption in this paper:

Assumption 5.2 *Under the model setting stated in Section 4.1, in this paper, we assume that* e_i *and* ε_j *are independent for any* i *and* j *and we assume that* N *is very large (it can be million, billion, zillion, or much larger) but it is finite.*

We note that when the null hypothesis H_0 holds, we have $E\left(e_i \varepsilon_j\right) = 0$ for any i and j. Thus, it is not unreasonable to assume e_i and ε_j are independent for any i and j. We also note that the reason we assume that N is very large (it can be million, billion, zillion, or much larger) but it is finite is to make the proof simple. If we do not assume this, the proof will be more complicated. To obtain the theory we developed in this paper, we first state the following theorem:

Theorem 5.3 *If* X_t *is defined by Equation* (4.2)*, then* $E(s_X^2) \to \infty$ *and* $N \to \infty$ *where* s_X^2 *is defined by* (3.4)*.*

Proof. Refer to Wong and Pham (2025). ■

Corollary 5.1 *If* X_t *is defined by Equation* (4.2)*, then* $Var(s_X^2) \to \infty$ *and* $N \to \infty$*.*

Proof. Refer to Wong and Pham (2025). ■

Now, we turn to develop the main theorem in this paper as follows:

Theorem 5.4 *Under the model setting stated in Section 4.1, let* $\mathbf{T}_N$ *be defined in Definition 5.1. if Assumption 5.2 holds, then* $E(\mathbf{T}_N) \neq 0$*.*

Proof. Refer to Wong and Pham (2025). ■

Theorem 5.5 *Under the model setting stated in Section 4.1, let* $\mathbf{T}_N$ *be defined in Definition 5.1. if Assumption 5.2 holds, then* $E(\mathbf{T}_N^2) \neq 1$.

Proof. Refer to Wong and Pham (2025). ■

In the above proof, we assume Assumption 5.2 holds so that N is very large (and it can be million, billion, zillion or much larger) but it is finite so that we have

$$\sum_{t=1}^{N}\left(\sum_{j=1}^{t}\frac{1-\phi^{2t}}{1-\phi^2}\right) \to \frac{1}{1-\phi^2} \times \frac{N(N+1)}{2} \; . \tag{5.2}$$

We note that if Assumption 5.2 does not hold in the sense that N does tend to infinite, then Equation (5.2) does not hold.

Problem 5.1 *If Assumption 5.2 does not hold in the sense that* N *does tend to infinite, then Equation* (5.2) *does not hold such that*

$$\sum_{t=1}^{N}\left(\sum_{j=1}^{t}\frac{1-\phi^{2t}}{1-\phi^2}\right) \neq \frac{1}{1-\phi^2} \times \frac{N(N+1)}{2} \; . \tag{5.3}$$

We post Problem 5.1 in this paper as an open problem to the literature, see if any academic can get the answer for $\sum_{t=1}^{N}\left(\sum_{j=1}^{t}\frac{1-\phi^{2t}}{1-\phi^2}\right)$.

We also note that if Assumption 5.2 does not hold in the sense that N tends to infinity, then the assertion of Theorem 5.5 still holds, but the proof is more complicated, so we skip discussing this situation in the paper.

Theorem 5.6 *Under the model setting stated in Section 4.1, if* $\mathbf{T}_N$ *is defined in Definition 5.1 and Assumption 5.2 holds, then* $\mathbf{T}_N$ *does not follow* t*-distribution with* $N-2$ *degrees of freedom, and thus, the traditional* T *test,* T_N*, shown in (3.7) cannot be used to test for the correlation between a stationary series and a series with a unit root.*

Proof. Firstly, from Theorem 5.5, we get $E(\mathbf{T}_N) \neq 0$ but from Theorem 5.1, we get $E(T_N) = 0$. So, $\mathbf{T}_N$ is different from T_N.

Second, from Theorem 5.5, we get $Var(\mathbf{T}_N) \neq 1$ but from Theorem 5.2, we get $Var(T_N) = 1$. So, $\mathbf{T}_N$ is different from T_N.

Third, from Proposition 3.1, we know that the numerator of T_N becomes Z which is distributed as $N(0, 1)$, the denominator of T_N is distributed as χ^2_{N-2}, and the former is independent of the latter so that T_N is distributed as $T_{(N-2)}$. However, looking into the Equation (**??**), one will know that the numerator of $\mathbf{T}_N$ is so complicated that it cannot be distributed as $N(0, 1)$, the denominator of $\mathbf{T}_N$ is also so complicated that it cannot be distributed as χ^2_{N-2}, and it is impossible for the former to be independent of the latter so that $\mathbf{T}_N$ cannot be distributed as $T_{(N-2)}$. For example, the numerator of $\mathbf{T}_N$ is the square root of

$$\frac{\left\{\sum_{t=1}^{N}\left[\left(\sum_{i=1}^{t} e_i\right)\left(\sum_{j=1}^{t} \phi^{t-j}\varepsilon_j\right)\right]\right\}^2}{\sum_{t=1}^{N}\sum_{i=1}^{t} e_i^2 \times \sum_{t=1}^{N}\left(\sum_{j=1}^{t} \phi^{t-j}\varepsilon_j\right)^2}, \tag{5.4}$$

which contain the square root of the components containing

$$\frac{\left\{\sum_{t=1}^{N}\left[\left(\sum_{i=1}^{t} Z\right)\left(\sum_{j=1}^{t} \phi^{t-j} Z\right)\right]\right\}^2}{\sum_{t=1}^{N}\sum_{i=1}^{t} U \times \sum_{t=1}^{N}\left(\sum_{j=1}^{t} \phi^{2(t-j)} U\right) + K}, \tag{5.5}$$

where $Z \sim N(0, 1)$ and $U \sim \chi^2_{(1)}$. Thus, it cannot be distributed as $N(0, 1)$.

Based on the above discussion, we conclude that $\mathbf{T}_N$ does not follow t-distribution with $N-2$ degrees of freedom, and thus, the traditional T test, T_N, shown in (3.7) cannot be used to test for the correlation between a stationary series and a series with a unit root. ■

We note that the results from Table 4.1 confirm that when Y_t follows a stationary AR(1) model and X_t follows a random-walk model, then the correlation of a Y_t and X_t may not be able to get any meaningful outcome, the correlation of a Y_t and X_t could be spurious, and the test, T_N, from the standard correlation may not be able to be used to test a correlation of Y_t and X_t when Y_t follows a stationary AR(1) model and X_t follows a random-walk model.

6 Conclusion and Future Study

We have reviewed several papers in the literature and found that some papers report correlations between a stationary Y_t and a non-stationary X_t. For example, Singh et al. (2011) found that GDP (which is non-stationary) has a positive relationship with stock returns (which are stationary). Other papers study correlations between GDP (non-stationary) and various potentially stationary variables, such as those by Lakstutiene (2008), Vinkler (2008), Andrei, *et al.* (2009), Valadez (2011), Szigeti, *et al.* (2013), Marcu, *et al.* (2015), Bilyuga, *et al.* (2016), Anghelache, *et al.* (2019), and others. However, to our knowledge, no study directly addresses whether the traditional correlation test is appropriate for testing the correlation between a stationary time series and a non-stationary time series. To bridge this gap, in this paper, we first conjecture that the correlation between a stationary Y_t and a non-stationary X_t could be spurious if standard correlation tests are used, and that such a correlation may not yield any meaningful results. For comparison, we also conjecture that traditional correlation tests can be applied to the correlations between two random series Y_t and X_t. At last, we set the conjecture to see whether the distributions of the test statistics are different for the two cases.

In the simulations, we model Y_t using a stationary AR(1) model and X_t using a non-stationary random walk model. We then simulate 10,000 pairs of X_t and Y_t for each combination of the AR parameter ϕ and sample size N. For each pair, we calculate the traditional correlation coefficient $\hat{\rho}_{xy}$ and use the t-test to obtain the statistic and its corresponding p-value. The p-values are compared to 0.05, and we compute the proportion of p-values smaller than 0.05, which we refer to as the "rejection rate". We conduct similar simulations for two random series Y_t and X_t.

Our simulations reveal several findings. First, we find that when Y_t follows a stationary AR(1) model with AR parameter ϕ and X_t follows a random-walk model, the rejection rate exceeds 5%, increasing as ϕ rises (from around 7.86% at $\phi = 0.1$ to around 62.3% at $\phi = 0.9$). This confirms the first three conjectures and suggests that when Y_t is stationary and X_t is non-stationary, the correlation between Y_t and X_t may not yield any meaningful result and it could be spurious. Moreover, we confirm that standard correlation tests are not appropriate. On the other hand, we find that when Y_t is a random series and X_t follows a random-walk model, the rejection rate is close to 5%, implying that the conjectures could still hold and that the correlation between Y_t and X_t is not spurious and could yield meaningful results, and the standard correlation tests could still be used to test whether the correlation between Y_t and X_t is zero. Last, we find that when both Y_t and X_t follow random-walk models, the rejection rate averages 86.29%, much higher than 5%. This suggests that the correlation between two random-walk series may not yield any meaningful result, the correlation could be spurious, and the standard correlation tests may not be appropriate for testing the correlation between Y_t and X_t.

Our simulations reveal several findings. First, we find that when Y_t follows a stationary AR(1) model with AR parameter ϕ and X_t follows a random-walk model, the rejection rate exceeds 5%, increasing as ϕ rises (from around 7.86% at $\phi = 0.1$ to around 62.3% at $\phi = 0.9$). This confirms the first three conjectures and suggests that when Y_t is stationary and X_t is non-stationary, the correlation between Y_t and X_t may not yield any meaningful result and it could be spurious. Moreover, we confirm that standard correlation tests are not appropriate. On the other hand, we find that when Y_t is a random series

and X_t follows a random-walk model, the rejection rate is close to 5%, implying that the conjectures could still hold and that the correlation between Y_t and X_t is not spurious and could yield meaningful results, and the standard correlation tests could still be used to test whether the correlation between Y_t and X_t is zero. Last, we find that when both Y_t and X_t follow random-walk models, the rejection rate averages 86.29%, much higher than 5%. This suggests that the correlation between two random-walk series may not yield any meaningful result, the correlation could be spurious, and the standard correlation tests may not be appropriate for testing the correlation between Y_t and X_t.

We then develop an estimation and testing theory for the correlation between a stationary Y_t and a non-stationary X_t and prove that the standard correlation test cannot be used in this case. To do so, we first show that if Y_t and X_t are random, then $E(T_N) = 0$ and $E\left(T_N^2\right) \to 1$ as $N \to \infty$. On the other hand, we prove that when Y_t follows a stationary AR1 model and X_t is $I(1)$, then $E(T_N) \neq 0$ and $E\left(T_N^2\right) \neq 1$ where T_N is the standard correlation test. Moreover, it is well-known that the standard correlation test T_N is distributed as the t-distribution with $N-2$ degrees of freedom if $H_0 : \rho = 0$ holds. However, we find that even $H_0 : \rho = 0$ holds, when Y_t follows a stationary AR1 model and X_t is $I(1)$, then the standard correlation test T_N is not distributed as the t-distribution with $N-2$ degrees of freedom. Thus, we confirm that the standard correlation test cannot be used for the case when Y_t follows a stationary AR1 model and X_t is $I(1)$.

We note that our simulations demonstrate three distinct scenarios: (1) when Y_t follows a stationary AR(1) model and X_t a random-walk model, (2) when Y_t is a random series and X_t a random-walk model, and (3) when both Y_t and X_t follow random-walk models. However, in this paper, we only develop the estimation and testing theory for the first case. Further studies could extend the theory to the second and third cases. Moreover, future research could explore other scenarios, such as when the stationary series follows a different process. In addition, researchers could investigate remedial methods for handling these situations. Since many empirical studies rely on correlations between stationary and non-stationary time series, further work should extend the literature to apply the correct testing methodology in such cases. Moreover, since the model used in our paper is very simple, extensions of our paper include studying more complicated models or studying some related areas, see, for example, Wong and Pham (2022), Wong and Pham (2023), Wong and Pham (2024), and others.

References

Abeysinghe, T. (1994), 'Deterministic seasonal models and spurious regressions', *Journal of Econometrics* **61**(2), 259–272.

Agiakloglou, C. (2013), 'Resolving spurious regressions and serially correlated errors', *Empirical Economics* **45**(3), 1361–1366.

Brockwell, P. J. and Davis, R. A. (2002), *Introduction to time series and forecasting*, Springer.

Cheng, Y., Hui, Y., Liu, S. and Wong, W.-K. (2022), 'Could significant regression be treated as insignificant: An anomaly in statistics?', *Communications in Statistics: Case Studies, Data Analysis and Applications* **8**(1), 133–151.

Cheng, Y., Hui, Y., McAleer, M. and Wong, W.-K. (2021), 'Spurious relationships for nearly non-stationary series', *Journal of Risk and Financial Management* **14**(8), 366.

Engle, R. F. and Granger, C. W. (1987), 'Co-integration and error correction:

representation, estimation, and testing', *Econometrica: journal of the Econometric Society* pp. 251–276.

Granger, C. W. and Newbold, P. (1974), 'Spurious regressions in econometrics', *Journal of econometrics* **2**(2), 111–120.

Kao, C. (1999), 'Spurious regression and residual-based tests for cointegration in panel data', *Journal of econometrics* **90**(1), 1–44.

Kim, T.-H., Lee, Y.-S. and Newbold, P. (2004), 'Spurious regressions with stationary processes around linear trends', *Economics Letters* **83**(2), 257–262.

Marmol, F. (1995), 'Spurious regressions between i (d) processes', *Journal of Time Series Analysis* **16**(3), 313–321.

Nakatani, T. and Teräsvirta, T. (2009), 'Testing for volatility interactions in the constant conditional correlation garch model', *The Econometrics Journal* **12**(1), 147–163.

Pesaran, M. H., Shin, Y. and Smith, R. P. (1999), 'Pooled mean group estimation of dynamic heterogeneous panels', *Journal of the American statistical Association* **94**(446), 621–634.

Phillips, P. C. (1986), 'Understanding spurious regressions in econometrics', *Journal of econometrics* **33**(3), 311–340.

Singh, T., Mehta, S. and Varsha, M. (2011), 'Macroeconomic factors and stock returns: Evidence from taiwan', *Journal of economics and international finance* **3**(4), 217.

Sun, Y. (2004), 'A convergent t-statistic in spurious regressions', *Econometric Theory* **20**(5), 943–962.

Tsay, R. S. (1989), 'Testing and modeling threshold autoregressive processes', *Journal of the American statistical association* **84**(405), 231–240.

Tsay, W.-J. and Chung, C.-F. (2000), 'The spurious regression of fractionally integrated processes', *Journal of Econometrics* **96**(1), 155–182.

Ventosa-Santaulária, D. (2009), 'Spurious regression', *Journal of Probability and Statistics* **2009**.

Westerlund, J. (2008), 'Panel cointegration tests of the fisher effect', *Journal of Applied Econometrics* **23**(2), 193–233.

Wong, W.-K., Cheng, Y. and Yue, M. (2024), 'Could regression of stationary series be spurious?', *Asia-Pacific Journal of Operational Research* .

Wong, W.-K. and Pham, M. T. (2022), 'Could the test from the standard regression model could make significant regression with autoregressive noise become insignificant?', *The International Journal of Finance* **34**, 1–18.

Wong, W.-K. and Pham, M. T. (2023), 'Could the test from the standard regression model could make significant regression with autoregressive $\mathrm{y_t}$ and $\mathrm{x_t}$ become insignificant?', *The International Journal of Finance* **35**, 1–19.

Wong, W.-K. and Pham, M. T. (2024), 'How to model a simple stationary series with a non-stationary series?', *The International Journal of Finance* . forthcoming.

Wong, W.-K. and Pham, M. T. (2025), 'Could the correlation of a stationary series with a non-stationary series obtain meaningful outcomes?', *Annals of Financial Economics* . Forthcoming.

Wong, W.-K. and Yue, M. (2024), 'Could regressing a stationary series on a non-stationary series obtain meaningful outcomes?', *Annals of Financial Economics* **19**(03), 2450011.