

AN ALTERNATIVE APPROACH FOR TESTING FOR LINEAR ASSOCIATION FOR TWO INDEPENDENT STATIONARY AR(1) PROCESSES

Christos Agiakloglou

► To cite this version:

Christos Agiakloglou. AN ALTERNATIVE APPROACH FOR TESTING FOR LINEAR ASSOCIA-TION FOR TWO INDEPENDENT STATIONARY AR(1) PROCESSES. Applied Economics, 2011, 44 (36), pp.4799-4803. 10.1080/00036846.2011.595695 . hal-00730233

HAL Id: hal-00730233 https://hal.science/hal-00730233

Submitted on 8 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Submitted Manuscript



AN ALTERNATIVE APPROACH FOR TESTING FOR LINEAR ASSOCIATION FOR TWO INDEPENDENT STATIONARY AR(1) PROCESSES

Journal:	Applied Economics		
Manuscript ID:	APE-2010-0484.R1		
Journal Selection:	Applied Economics		
Date Submitted by the Author:	01-Jun-2011		
Complete List of Authors:	Agiakloglou, Christos; University of Piraeus, Economics		
JEL Code:	C22 - Time-Series Models < C2 - Econometric Methods: Single Equation Models < C - Mathematical and Quantitative Methods, C10 - General < C1 - Econometric and Statistical Methods: General < C - Mathematical and Quantitative Methods		
Keywords:	Correlation coefficient , Spurious correlations , Stationary AR(1) processes , Test for linear association		

SCHOLARONE[™] Manuscripts

Editorial Office, Dept of Economics, Warwick University, Coventry CV4 7AL, UK

AN ALTERNATIVE APPROACH FOR TESTING FOR LINEAR ASSOCIATION FOR

TWO INDEPENDENT STATIONARY AR(1) PROCESSES

Christos Agiakloglou* and Apostolos Tsimpanos

University of Piraeus, Greece

Abstract

Spurious correlations occur when two independent time series are found to be correlated according to the typical statistical procedure for testing the null hypothesis of zero correlation in the population. Using a Monte Carlo analysis, this study examines the spurious correlation phenomenon for two independent stationary AR(1) processes and it finds that if an alternative testing procedure is applied, spurious behavior is eliminated using the variance of the sample correlation coefficient of these two series, suggested by Bartlett (1935).

Keywords: Correlation coefficient, spurious correlations, stationary AR(1) processes, testing for linear association

JEL Classification: C22

 Corresponding address: *Department of Economics, University of Piraeus, Greece Karaoli & Dimitriou 80, Piraeus 18534, Greece. Tel.: +30-210-414-2290, Fax.: +30-210-414-2290 E-mail address: agiaklis@unipi.gr

1. Introduction

In many economic applications interest is focused on investigating the existence of a linear relationship between two random variables using the sample correlation coefficient which declares the strength of the linear association in the sample and can be used as the basis of testing the null hypothesis of no linear correlation in the population. In some cases, however, an analyst may get misleading statistical results with respect to the degree and to the existence of a linear association between two random variables, when indeed these two random variables are not linearly related. This phenomenon is known as the spurious correlation phenomenon, introduced by Yule (1926), although in the literature it received a considerable attention as the spurious regression phenomenon presented by the pioneer work of Granger and Newbold (1974), in which, using a Monte Carlo analysis, they showed that the regression of two independent random walk processes generates spurious results. Furthermore, the behavior of this phenomenon was well-documented mathematically by Phillips (1986) and extended by Granger *et al* (2001) for two stationary independent AR(1) processes.

Spurious correlations and spurious regressions are two similar if not identical terms referring to the same phenomenon of obtaining false evidence for the existence of a linear relationship between two variables. However, in the case of spurious correlations the analyst has no indication about the existence of such behavior, unless he or she has some *a priori* information about the relationship between these two variables, such that a high absolute value of the sample correlation coefficient will be considered suspicious, see for example Hendry (1980). Contrary, in the case of spurious regressions, Granger

Submitted Manuscript

and Newbold (1974) have pointed out that a low value of the Durbin-Watson statistic will appear in regression analysis and standard procedures, such as the Cochrane-Orcutt, will fail to correct the problem of autocorrelated errors. Actually, evidence of serially correlated errors will also appear in regression analysis, along with spurious results, even in the case of two independent stationary AR(1) processes, as Agiakloglou (2009) has shown. Thus, it is more difficult to detect spurious correlations than spurious regressions.

This paper demonstrates that using the variance of the sample correlation coefficient of two independent AR(1) processes, the spurious behavior is eliminated for all values of the autocorrelation coefficient. Recall, that Granger *et al* (2001) have also proposed a method to correct the problem of spurious regressions for stationary AR(1) processes, but their empirical results did not convert to nominal levels, especially for large values of the autoregressive parameter, even for large sample sizes. Only asymptotically Granger *et al* (2001) have shown convergence.

2. Testing for linear association

To investigate whether or not a linear association between two random variables exists, an analyst must test the null hypothesis of zero correlation, i.e., H_0 : $\rho = 0$, against the alternative of not zero correlation, i.e., H_1 : $\rho \neq 0$, where ρ is the population correlation coefficient. The test is implemented using the following t statistic:

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{T - 2}}}\tag{1}$$

Submitted Manuscript

where r is the sample correlation coefficient and T is the sample size. The t statistic follows a t distribution with (T - 2) degrees of freedom and the null hypothesis will be accepted if its absolute value is less than the critical value, indicating evidence of no linear association between the two variables. In the case of spurious correlations the null hypothesis will be rejected, although the two variables will not be linearly related.

Yule (1926) pointed out that frequently in practice a high degree of linear association between two independent time series can arise for variables that have nothing in common. As an effort to justify this behavior, Yule (1926) studied the properties of the sample correlation coefficient of two series and noticed that the major factor that determines this spurious behavior is the shape of the frequency distribution of the correlation coefficient of these two series. Banerjee *et al.* (1993), using Monte Carlo analysis, examined the frequency distribution of the correlation coefficient for various orders of independent integrated time series verifying Yule's (1926) initial results. They concluded that if the two processes are non-stationary I(1) processes, the frequency distribution has a U shape with values of -1 and +1 to be more likely to occur. Contrary, in the case where the two series are stationary white noise processes, the frequency distribution of the correlation coefficient will be symmetric around zero and it will look like normal distribution.

However, simulation results for two independent I(1) processes based on 10,000 replications will show that the standard deviation of the t statistic ranges from 7.38 to 23.37 for series of 100 to 1,000 observations, whereas the null hypothesis of zero correlation is rejected from 76.95% to 92.76% respectively at the 5% nominal level.

Things are even worse for two independent I(2) processes in which case the value of the standard deviation of the t statistic ranges from 49.01 to 139.39 and the null hypothesis is rejected from 94.82% to 98.18%, using the same simulation process as above. Therefore, the large number of rejections of the null hypothesis of zero correlation for non-stationary processes is basically due to the fact that the variance of the sample correlation coefficient is not properly computed and indeed its value strongly deviates from one as appose to the white noise case which remains one, regardless of the sample size.

Spurious correlations can also appear in the case of two independent stationary AR(1) processes, as indicated by Granger *et al* (2001) in the context of spurious regressions. To illustrate, consider two independent AR(1) processes X_t and Y_t generated from the following DGP:

$$X_t = \varphi_x X_{t-1} + \varepsilon_{xt}$$
 and $Y_t = \varphi_y Y_{t-1} + \varepsilon_{yt}$ (2)

where the errors ε_{xt} and ε_{yt} are white noise N(0, 1) processes independent of each other and the autoregressive parameters are allowed to take values of 0.0, 0.2, 0.5, 0.8 and 0.9.

Table 1 reports the percentage of rejections of the null hypothesis of zero correlation against a two sided alternative at the 5% nominal level for two independent stationary AR(1) processes for sample sizes of 100, 500 and 1,000 observations based on 10,000 replications. The most interesting feature of this table is the fact that, unlike the two non-stationary cases, previously discussed and especially the I(1) case, the percentage of rejections of the null hypothesis of zero correlation remains unchanged regardless of the sample size and it is only affected by the magnitude of the autoregressive parameter. Thus, one will get more spurious results as the value of the autoregressive parameter increases. For example, for $\varphi_x = \varphi_y = 0.5$, the null hypothesis is

Submitted Manuscript

rejected approximately 13% at the 5% nominal level, whereas for $\varphi_x = \varphi_y = 0.9$, this number becomes approximately 52%.

However, following Yule's (1926) proposition, an analyst will find no spurious results based on the frequency distribution of the correlation coefficient of two independent stationary AR(1) processes. This distribution remains symmetric around zero mean regardless of the value of the autoregressive parameter, as, for example, Figures 1 and 2 depict. Therefore, as in the case of non-stationary processes, spurious correlations appear simply because the value of the standard deviation of the t statistic for testing the null hypothesis of zero correlation is not equal to one for all values of the autoregressive parameter. Indeed, as Table 2 reports, this value is only affected by the magnitude of the autoregressive parameter of the series and not by the sample size. For example, the standard deviation of the t statistic for $\varphi_x = \varphi_y = 0.5$ is close to 1.3, whereas for $\varphi_x = \varphi_y = 0.9$ this number is approximately equal to 3. Hence, although, the frequency distribution for the t statistic is symmetric around zero mean, it becomes flatter than the standard normal distribution as the value of the autoregressive parameter increases, as, for example, can be seen from Figures 3 and 4.

3. Simulation results using the variance of the sample correlation coefficient of two independent stationary AR(1) processes

Having discussed the problem of spurious correlations for two independent stationary AR(1) processes the next step is to investigate the behavior of this phenomenon using the variance of their sample correlation coefficient obtained as follows. For two independent

stationary AR(1) processes X_t and Y_t with autocorrelation coefficients ρ_x and ρ_y respectively we have:

 $E[\frac{1}{T^{2}}(\sum X_{t}Y_{t})^{2}] = \frac{1}{T^{2}}E[\sum X_{t}^{2}Y_{t}^{2} + 2\sum_{t \neq s}X_{t}Y_{t}X_{s}Y_{s}]$ (3)

and since

$$\sum_{t \neq s} X_t Y_t X_s Y_s = \sum X_t X_{t+1} Y_t Y_{t+1} + \sum X_t X_{t+2} Y_t Y_{t+2} + \dots$$

equation (3) is written as:

$$E[\frac{1}{T^2}(\sum X_t Y_t)^2] = \frac{1}{T}\sigma_x^2 \sigma_y^2 + \frac{2\sigma_x^2 \sigma_y^2}{T^2}[(T-1)\rho_x \rho_y + (T-2)\rho_x^2 \rho_y^2 + \dots]$$

which approximately is equal to:

$$E[\frac{1}{T^2}(\sum X_t Y_t)^2] = \frac{1}{T}\sigma_x^2 \sigma_y^2 (1 + \frac{2\rho_x \rho_y}{1 - \rho_x \rho_y})$$

Hence, the variance of the sample correlation coefficient between the two independent stationary AR(1) series is approximately defined as:

$$Var(r) = \frac{1}{T} \left(\frac{1 + \rho_x \rho_y}{1 - \rho_x \rho_y} \right)$$

or equivalently as:

$$Var(r) = \frac{1}{T} \left(\frac{1 + \varphi_x \varphi_y}{1 - \varphi_x \varphi_y} \right)$$
(4)

since $\rho_x = \varphi_x$ and $\rho_y = \varphi_y$ for AR(1) processes. More evidence about the proof of this variance can be found in Bartlett (1935), whereas McGregor (1962) verifies the existence of this variance by determining the approximate null distribution of the sample correlation coefficient of two stationary Markov chain processes using the steepest descents method proposed by Daniels (1954 and 1956). Clearly, the degree of accuracy of this variance depends on the sign, on the absolute magnitude of the two autoregressive

Submitted Manuscript

coefficients and on the sample size. One should expect less accuracy if φ_x and φ_y are both positive (or negative), if their absolute magnitude is close to one, as well as if the sample size is small.

To investigate the accuracy of this variance in the context of spurious correlations two independent AR(1) processes X_t and Y_t are generated by the same DGP, given in (2). Based on the sample correlation coefficient of these two series, the test for zero correlation is conducted by replacing the denominator of the t statistic defined in equation (1) with the square root of the variance of equation (4) and the results of 10,000 replications at the 5% nominal level are reported on Table 3. Perhaps the most astonishing feature of this table is the fact that for all values of the autoregressive parameter the empirical levels are very close to nominal levels, due to the fact that the value of the standard deviation of the t statistic for testing the null hypothesis of zero correlation is one for all cases. Deviation from this behavior we only observe for values of the autoregressive parameter of 0.8 and 0.9 and only for series of sample size of 100 observations in which cases the empirical levels are smaller than the nominal ones, since the standard deviation of the t statistic in these cases was slightly smaller than one, i.e., its values ranged from 0.88 to 0.95.

4. Concluding remarks

Spurious correlations can be found not only for a pair of two independent non-stationary processes, but also for a pair of two independent stationary AR(1) processes. However, using the approximate variance of the sample correlation coefficient of two independent stationary AR(1) processes, this study shows that the spurious behavior can be eliminated

for all values of the autoregressive parameter. Perhaps, the most interesting feature of his study is the fact that the correction of spurious results can be obtained very easily. Indeed, as long as two processes are identified as AR(1) processes, using the Box and Jenkins (1976) methodology, the estimates of the autoregressive parameters can be used to determine the value of the variance of the sample correlation coefficient in order to apply the test for a linear association.

References

- Agiakloglou, C. (2009), Evidence of ARCH(1) errors in the context of spurious regressions, *Communications in Statistics Simulations and Computation*, **38**, 1803-1810.
- Banerjee, A., Dolado, J., Galbraith, J. W. and Hendry, D. F. (1993). Co-integration, error correction and the econometric analysis of non-stationary data, Oxford University Press, Oxford.
- Bartlett, M. S. (1935). Some aspects of the time-correlation problems in regard to tests of significance, *Journal of the Royal Statistical Society*, **98**, 536-543.
- Box, G. E. P. and G. M. Jenkins, (1976), Time Series Analysis Forecasting and Control, Holden-Day, Oakland.
- Daniels, H. E. (1954). Saddlepoint approximations in statistics, Ann. Math Statistics, 25, 631-650.
- Daniels, H. E. (1956). The approximate distribution of serial correlation coefficients, *Biometrika*, **43**, 169-185.
- Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics, *Journal* of *Econometrics*, **2**, 111-120.
- Granger, C. W. J., Hyung, N. and Jeon, Y. (2001). Spurious regressions with stationary series, *Applied Economics*, **33**, 899-904.
- Hendry, D. F. (1980). Econometrics Alchemy or science?, *Economica*, 47, 387-406.
- McGregor, J. R. (1962). The approximate distribution of the correlation between two stationary linear Markov series, *Biometrika*, **49**, 379-388.
- Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics, *Journal of Econometrics*, **33**, 311-340.
- Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between timeseries? A study in sampling and the nature of time-series, *Journal of the Royal Statistical Society*, , 1-64.

1	
1	
2	
3	
4	
5	
6	
7	
1	
8	
9	
1	0
1	1
4	י ר
	2
1;	3
1	4
1	5
1	6
1	7
1	, 0
10	0
19	9
2	0
2	1
2	2
2	2
2	د ۸
2	4
2	5
2	6
2	7
2	R
2	0
23	9
3	0
3	1
3	2
3	3
2	1
3	4
3	5
3	6
3	7
3	8
2	n n
3	9
4	U
4	1
42	2
4	3
۵.	4
л. Л	5
4	5
4	b
4	7
4	8
4	9
5	ñ
ט ר	4
5	1
5	2
5	3
5	4
5	5
5	6
	7
S	<i>(</i>
5	8
5	9
6	0

Table 1
Percentage of rejections of the null hypothesis of zero correlation at the 5% nominal
level (t > 1.96) for two independent stationary AR(1) processes based on 10,000
replications

			1			
		φ_x				
Т	φ_y	0.0	0.2	0.5	0.8	0.9
	0.0	5.03				
	0.2	5.18	6.20			
100	0.5	5.38	7.77	12.85		
	0.8	5.04	9.86	19.53	35.23	
	0.9	4.81	10.44	22.42	41.12	50.27
	0.0	5.24				
	0.2	5.13	5.77			
500	0.5	5.31	7.92	13.15		
	0.8	4.90	9.44	19.72	34.78	
	0.9	5.42	10.46	22.67	42.56	52.22
	0.0	5.10				
	0.2	5.25	6.00			
1,000	0.5	4.73	7.38	12.87		
	0.8	5.30	9.94	19.92	36.05	
	0.9	4.83	10.53	22.55	43.21	52.01

Table 2

Standard deviation of the t statistic for testing the null hypothesis of zero correlation for two independent stationary AR(1) processes based on 10,000 replications

		φ _x				
Т	φ_y	0.0	0.2	0.5	0.8	0.9
	0.0	1.00				
	0.2	1.02	1.05			
100	0.5	1.01	1.11	1.30		
	0.8	1.00	1.17	1.52	2.12	
	0.9	1.02	1.21	1.62	2.45	3.00
	0.0	1.00				
	0.2	1.01	1.04			
500	0.5	1.00	1.11	1.29		
	0.8	1.02	1.18	1.51	2.10	
	0.9	1.00	1.21	1.62	2.48	3.10
	0.0	1.00				
	0.2	1.00	1.05			
1,000	0.5	1.00	1.11	1.28		
	0.8	1.01	1.17	1.52	2.15	
	0.9	1.00	1.20	1.61	2.47	3.07

Percentage of rejections of the null hypothesis of zero correlation at the 5% nominal level (|t| > 1.96) for two independent stationary AR(1) processes using the approximate variance of their sample correlation coefficient based on 10,000 replications

		φ_x				
Т	φ_y	0.0	0.2	0.5	0.8	0.9
	0.0	4.76				
	0.2	5.22	5.13			
100	0.5	5.18	5.08	4.56		
	0.8	4.96	4.72	4.59	3.41	
	0.9	5.08	4.73	4.46	2.75	1.59
	0.0	5.21				
	0.2	5.09	5.03			
500	0.5	5.07	5.03	5.25		
	0.8	4.91	4.77	5.07	4.81	
	0.9	5.09	4.88	5.11	4.61	4.56
	0.0	4.98				
	0.2	5.21	4.81			
1,000	0.5	4.90	5.13	4.90		
	0.8	4.80	5.08	5.09	4.77	
	0.9	5.09	4.88	4.92	4.79	5.16

