Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v7n1p1

The Modified R a Robust Measure of Association for Time Series

By Rehman and Malik

April 26, 2014

Electronic Journal of Applied Statistical Analysis
Vol. 7, Issue 1, 2014, 1-13
DOI: 10.1285/i20705948v7n1p1

The Modified R a Robust Measure of Association for Time Series

Atiq-ur-Rehman* and Muhammad Irfan Malik

International Institute of Islamic Economics, International Islamic University Islamabad, Pakistan

April 26, 2014

Since times of Yule (1926), it is known that correlation between two time series can produce spurious results. Granger and Newbold (1974) see the roots of spurious correlation in non-stationarity of the time series. However the study of Granger IV et al. (2001) prove that spurious correlation also exists in stationary time series. These facts make the correlation coefficient an unreliable measure of association. This paper proposes 'Modified R' as an alternate measure of association for the time series. The Modified R is robust to the type of stationarity and type of deterministic part in the time series. The performance Modified R is illustrated via extensive Monte Carlo Experiments.

keywords: Correlation Coefficient, Spurious Regression, Stationary Series.

1 Introduction

The correlation coefficient is one of most commonly used statistics in the analysis of statistical data. The correlation gives the measure of degree of association between two variables. The correlation coefficient was introduced by Pearson (1894) to analyze the cross-sectional data. This simple statistics is useful to have an immediate idea of the strength of relation between two variables, however it use for the time series data has been subject to critique since Yule (1926). The reason for this critique is the possibility of spurious correlation. The nature of time series is different from the cross sectional data. Most important feature of the time series data is the serial dependence of the observations. Due to the serial dependence, distribution of various estimators differs for

*Corresponding Author: ateeqmzd@gmail.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

time series than from the cross section. This serial dependence is one of the sources of spurious correlation among two data sets.

Concept of spurious correlation is as old in statistics as the correlation itself. Spurious correlations occurs if two variables appear to be correlated (or highly correlated) but in fact there is no (or very weak) relationship between the two variables. Pearson is responsible for the introduction of concept of spurious correlation. Aldrich (1995) reports his first encounter of Pearson with the spurious correlation during a study of personal judgment in 1897. Elderton (2011) published a book entitled 'Frequency-Curve and Correlation', which is assumed to be representative of Pearsonian School of Thought. In his book, he writes:

> *It is possible to obtain significant value of correlation coefficient when in reality two functions are absolutely uncorrelated (cited in Aldrich (1995)).*

Various types of occurrences of spurious correlation has been discovered in the first three decades of twentieth century e.g. (i) spurious correlation due to use of ratios (Yule, 1897), (ii) spurious correlation due to mixing of races (Pearson et al., 1899), Yule's illusory association that does not indicate causal relation but common dependence on third variable (Yule, 1910, 1926; Simon, 1954), and Yule's none-sense regression of time series. However now a days, the term 'spurious correlation' is used only in the context of time series. For example, Granger IV et al. (2001) (GHJ01 hereafter) define spurious correlation as:

> *A spurious regression occurs when a pair of independent series, but with strong temporal properties, is found apparently to be related according to standard inference in an OLS regression.*

Therefore the use of term spurious correlation is used only in the context of time series data. This study introduces a new and robust measure of association for the time series. This measure of association is not affected by the serial correlation in the time series and gives a reliable measure of association.

2 Correlation and Non-Stationarity

Most important reason for the development of literature on unit root is perhaps the possibility of spurious correlation among time series. Although there may be various reasons for occurrence of spurious correlation, and stationary series may also exhibit spurious correlation (GHJ01), most of literature present unit root as only responsible for the existence of spurious correlation. The correlation coefficient fails to measure strength of association between two unit root series. The historical roots of this phenomenon dates back to Yule (1926). He found that two time series may have extraordinarily high correlation when in fact there is no relationship between two. However there was no clarity on the reasons behind occurrence of this phenomenon. Yule (1926) and Simon (1954) thought that it is some missing variable which is responsible for such extraordinarily high correlation.

Granger and Newbold (1974) (GN74 hereafter) observed that if two series containing unit root are regressed onto each other, they are most likely to produce high R-square and significant t-statistics. Nelson and Plosser (1982) observed that unit root model is more appropriate for most of economic time series. These two studies present an alternate explanation of the phenomena observed by Yule a half century ago i.e. it's the unit root responsible for existence of spurious correlation. Phillips (1987) provided analytical proofs of the occurrence of spurious correlation and other diagnostic for the integrated series.

Later on, it has been proven that two independent stationary series when regressed onto each other, may also produce spurious results (GHJ01), most of econometric textbooks present existence of unit root an analogue of the spurious regression (Greene, 2003; Gujarati and Porter, 1999; Enders, 2008) for example.

3 Correlation Coefficient in Time Series

The correlation coefficient was introduced to analyze the cross-sectional data. The nature of time series is different from the cross sectional data. Most important feature of the time series data is the serial dependence of the observations. Due to the serial dependence, distribution of various estimators differs for time series than from the cross section. This serial dependence is one of the sources of spurious correlation among two data sets.

The time series are usually classified into two classes, stationary and unit root. For the unit root series, GN74 have proved high probability of the spurious results. For stationary series, it was believed that the asymptotic distribution of the regression diagnostic is similar to the that of IID series, however, GHJ01 have proved that there is very high probability of spurious regression for the stationary time series as well.

In a letter study GHJ01 found that the phenomenon of spurious regression is not restricted to the non-stationary variables.GHJ01 found evidence of spurious correlation between positively autocorrelated stationary series. In particular, the distribution of correlation coefficient is very different for autocorrelated series than from IID series so that one cannot use correlation coefficient as a measure of association among the time series. R-square and the multiple R are valid measure of association only if the two series are IID.

The properties of measures of correlation e.g. the correlation coefficient and the t-statistics presented above prove that the correlation coefficient r is not a reliable measure of association for the time series. This study proposes a simple statistics modified R as an alternate measure of association between the time series. The distribution is robust to amount of autocorrelation, the specification of deterministic regressors and the type of stationarity among the time series. The performance of this statistics is illustrated for various types of models via Monte Carlo experiments.

4 Spurious Correlation among Stationary Time Series

Consider two stationary series:

$$x_t = a_1 + d_1 x_{t-1} + e_t \qquad e_t \sim IN(0, S_e^2) \tag{1}$$

$$y_t = a_2 + d_2 y_{t-1} + u_t \qquad u_t \sim IN(0, S_u^2) \tag{2}$$

If there is no correlation between x_t and y_t , then the two series are independent and any reasonable measure of association should reveal this independence. However, if is non-zero, than the distribution of correlation coefficient R has wider spread and the series appear to be dependent. This spurious correlation diminish very slowly with the increase in sample size. To illustrate this fact, we had conducted a Monte Carlo experiment in which we have simulated the distribution of R. The results are presented in Fig1.

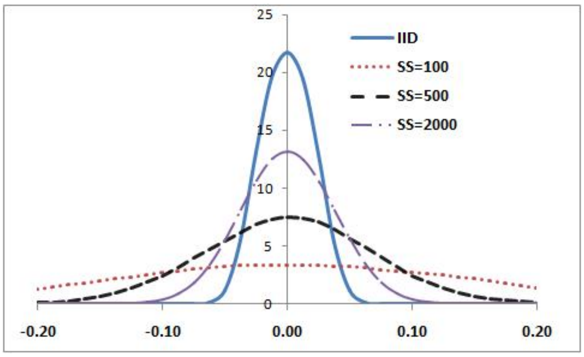

Figure 1: Distribution of Pearson correlation coefficient of stationary series with autocorrelation

The bold curve in Fig 1 represents the distribution of Pearson correlation R for IID series. This distribution is invariant to sample size. The dotted/dashed lines represent the distribution of R for different sample sizes when the two series are autocorrelated. The series have no mutually correlated, yet the spread of the distribution of R increased showing that the apparent correlation will be higher than the actual correlation. In Fig 1 it is visible that for sample size 2000, the distribution of correlation coefficient is much wider than the distribution of correlation for IID series. The 95%confidence interval for autocorrelation between two IID series is (0 +/- 0.103) when sample size is 2000. When series are autoregressive with autoregressive coefficient of 0.6, the correlation coefficient outside this interval was 55 %. If two series are generated by the DGP described in eq. 1 & 2 with $d_1 = d_2 = 0.9$, than the probability of correlation outside the interval increases to 77 % . This implies that even for the stationary series, the probability of spurious correlation may be as high as 77% with reasonably large sample size. With the increase in value of autoregressive parameter, the scatter of distribution of correlation increases and the probability of spurious correlation between series increases.

4.1 Implication of the above discussion

Although spurious correlation is usually referred to the unit root processes, one can see the incidence of extraordinarily high correlation in stationary series with autoregressive root not very close to unity. The distribution of correlation among autoregressive series converges to the distribution of correlation among IID series, but the convergence is very slow and in practically relevant sample sizes, there is huge probability of encounter with the spurious correlation among two uncorrelated time series. For unit root series, distribution of correlation among time series in not convergent and probability of spurious correlation increases with the increase of sample size.
All these problems make R-square an invalid inferential statistics whether the series are stationary or unit root. Therefore it is desirable to develop some statistics which can give more reliable measure of association between two time series.

5 Modified R

We propose a new statistics to be used as a measure of association between two time series. This statistics is the correlation between recursive forecasts errors of autoregressive models fitted to both the series.
Correlation coefficient is a descriptive assumed to measure strength of association between two data sets. When two data series are IID, correlation coefficient R gives reliable measure of the association, but when there is violation of IID property, the distribution of correlation coefficient changes drastically. We aim to develop a statistics whose distribution is invariant to the change of IID assumption. This can be done if we make a model which can approximate any of the dynamic structure. The forecast residual while forecasting from this type of model will be close to random. If we compute correlation between forecast residuals, it is expected to give more reliable measure of the association between the time series. Let us call this statistics as MR; the statistics is defined as follows:
For two time series of length T, let T1 be number smaller than T i.e. $T1 < T$ series $x = \{x_1, x_1.......x_T\}$ and $y = \{y_1, y_1.......y_T\}$

1. For $T1 < T$, estimate the autoregressive model $x_t = \widehat{a}_{T1} + \widehat{d}_{T1} x_{t-1} + e_t$ using OLS
2. compute $\widehat{x}_{T1+1} = \widehat{a}_{T1} + \widehat{b}_{T1} x_{T1}$
3. compute $e_{T1+1} = x_{T1+1} - \widehat{x}_{T1+1}$
4. Repeat the process for T1+1, T1+2, T-1 to compute $e_{T+1}, e_{T+2}, e_{T+3}.....e_T$
5. Repeat Steps 1-4 for the series for series $y = \{y_1, y_1.......y_T\}$ to compute forecast error $u_{T+1}, u_{T+2}, u_{T+3}.....u_T$
6. Compute the correlation between the forecast residuals from the two series of the forecast residuals

The modified R is based on recursive residuals from autoregressive model with generalized form of linear trend thus is capable of producing desired results. The MR is similar to Hough portmanteau statistics. However it differs from Haughs(1976) statistics in that Haughs statistics is based on residuals of ARMA model fitted to all points in the data sets; whereas MR is based on recursive residuals. It utilizes all past information to predict current residual of the time series and future observations are irrelevant for computation of current residual

6 Properties of Correlation Coefficient and the Modified R

We compute distribution of R and MR for data generated by following bivariate autoregressive process.

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} 1 \\ t \end{bmatrix} + \begin{bmatrix} \varepsilon_{xt} \\ \varepsilon_{yt} \end{bmatrix} \qquad \begin{bmatrix} \varepsilon_{xt} \\ \varepsilon_{yt} \end{bmatrix} \sim \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right]$$

For simplification of notation, we can rewrite as

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{Bd} + \boldsymbol{\varepsilon}_t \qquad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \tag{3}$$

This data generating process produces two autoregressive series. The two series will be independent if $\mathbf{\Sigma}$ is a diagonal matrix and correlated otherwise. So the value of any measure of association would depend only on $\mathbf{\Sigma}$. Other parameters in the model i.e. $\mathbf{A}$ and $\mathbf{b}$ are the nuisance parameters for the mutual correlation of two series. The correlation coefficient R has been shown to be dependent on the nuisance parameters. The matrix reflects the strength of autocorrelation in the two series whereas the matrix is to specify the deterministic part in the model. If $\mathbf{A} = \mathbf{0}$, this implies two series have no autocorrelation and $\mathbf{b} = \mathbf{0}$ implies absence of the drift and linear trend. Series would be IID if both $\mathbf{A}$ and $\mathbf{b}$ are zero. Non diagonal value of $\mathbf{\Sigma}$ implies by mutual correlation of the two series. We had Monte Carlo with different values of and results are summarized in Table 1 to 4. It was observed that MR is robust to the values of these nuisance parameters

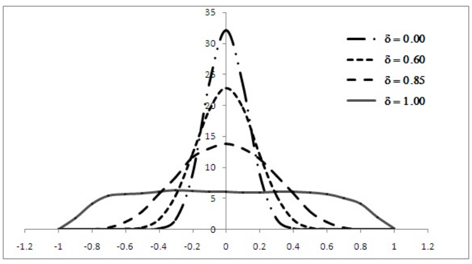

Figure 2: Distribution of R

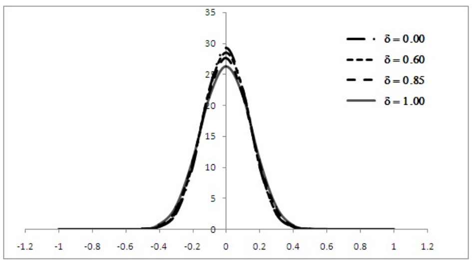

Figure 3: Distribution of MR

The fig. 1 & 2 above represent the distribution of R and MR where $\mathbf{b} = 0$ *and* $\mathbf{\Sigma} = \mathbf{I}$*. The strength of autocorrelation for two series remains same i.e* $d_1 = d_2 = d$*. We see that* *the distribution of R changes with the strength of autocorrelation whereas the distribution of MR is almost invariant to it.*

7 Testing the correlation

The correlation coefficient is used as a descriptive, not as a test of association. It gives quick idea of strength of correlation but neither used to prove nor to disprove the correlation between two variables. MR is supposed to serve the similar purpose, i.e. as a descriptive rather than a test. However, extraordinarily high correlation would lead someone to suspect that the two series are correlated. To measure the probability of deceivingly high correlation using the two measures of correlation, we used the two measures as a test for mutual correlation. Two sided critical values of R and MR were computed using two IID series. A value of R/MR outside these critical values is considered as rejection of null hypothesis of no correlation.

7.1 Results of Monte Carlo

By choosing various values of the nuisance parameters, we compute the distribution and rejection rate of null of no correlation under the null and alternative to compute the size and power respectively. The data generating process used for these simulations are described below Different models are constructed using different values of **b** and $\mathbf{\Sigma}$. The detail of these models is as under:

$M1 \quad \mathbf{b} = \mathbf{0} \quad and \quad \mathbf{\Sigma} = \mathbf{I} \quad (IID\ and\ mutually\ independent)$

$$M2 \quad \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \end{bmatrix} \quad and \quad \mathbf{\Sigma} = \mathbf{I} \quad (independent)$$

$$M3 \quad \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \quad and \quad \mathbf{\Sigma} = \mathbf{I} \quad (independent)$$

$$M4A \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad and \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} \quad (IID\ and\ weakly\ correlated)$$

$$M4B \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad and \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \quad (IID\ and\ strongly\ correlated)$$

$$M5A \quad \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \end{bmatrix} \quad and \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} \quad (weakly\ correlated)$$

$$M5B \quad \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \end{bmatrix} \quad and \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \quad (strongly\ correlated)$$

$$M6A \qquad \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \quad and \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} \qquad (weakly\ \ correlated)$$

$$M6B \qquad \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \end{bmatrix} \quad and \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \qquad (strongly\ \ correlated)$$

Table 1 summarize results of Monte Carlo when the two series do not have drift and linear trend. It is evident that the distribution of R is tremendously changed when the series are autocorrelated (not mutually correlated).
In all simulations reported in Table:1, only value of autocorrelation is changed and we can

Table 1: Distribution of R and MR for Different values of autoregressive parameter: DGP : M1, sample size 100

		Mean	Median	SD	IQR	RR
$d_x = 0.00$	R	0.00	0.00	0.12	0.16	5
$d_x = 0.00$	MR	0.00	0.00	0.13	0.18	5
$d_x = 0.60$	R	0.00	0.00	0.17	0.23	17
$d_y = 0.60$	MR	0.00	0.00	0.13	0.18	5
$d_x = 0.85$	R	0.00	0.00	0.27	0.39	41
$d_y = 0.85$	MR	0.00	0.00	0.14	0.19	7
$d_x = 1.00$	R	0.00	-0.01	0.49	0.82	72
$d_y = 1.00$	MR	0.00	0.00	0.15	0.20	8

see that due to this change, the distribution of R also changes but the distribution of MR remains same. The last row of the table corresponds to two mutually independent unit root series. The distribution of R and MR both are cantered at zero, however the spread of the distribution of R increase with increase in strength of autocorrelation. Hence the nuisance parameters enter the distribution of MR. But there is very little affect of nuisance parameters on the distribution of MR. The last column shows the rejection rate of null hypothesis of no autocorrelation. The deviation from IID assumptions leads the correlation to be biased to non-rejection of the relationship, whereas the MR maintains its size, so that the rejection of (true) null of no relationship does not deviate from its nominal size.
Table:2 summarizes the results of Monte Carlo when the two series contain drift but no linear trend. The chances of encounter with spurious relationship between two series increase when there is drift term present in the data generating process, if the relationship is measured by R. But the MR maintains its size even in the presence of drift term. Table:3 summarize results when the two series have both drift and linear trend. In that case R is almost sure to reject the null of no correlation.
Table:4-6 summarize results when the two series in data generating process have different strength of autocorrelation. First case in the three Tables corresponds to a situation when one series is integrated whereas other series is stationary. One can see that the performance of MR is affected by the presence of unit root series; however the size distortion is not very large. The superiority of MR over R is clearly evident.

Table 2: Distribution of R and MR for Different values of autoregressive parameter: DGP: M2

		Mean	Median	SD	IQR	RR
$d_x = 0.00$	R	0.00	0.00	0.12	0.16	5
$d_y = 0.00$	MR	0.00	0.00	0.13	0.17	5
$d_x = 0.60$	R	0.02	0.02	0.17	0.23	17
$d_y = 0.60$	MR	0.00	0.00	0.13	0.18	5
$d_x = 0.85$	R	0.14	0.15	0.25	0.36	45
$d_y = 0.85$	MR	0.00	0.00	0.13	0.19	6
$d_x = 1.00$	R	0.94	0.96	0.05	0.04	99
$d_y = 1.00$	MR	0.01	0.01	0.14	0.20	8

Table 3: : Distribution of R and MR for Different values of autoregressive parameter: DGP: M3

		Mean	Median	SD	IQR	RR
$d_x = 0.00$	R	0.99	0.99	0.00	0.00	100
$d_y = 0.00$	MR	0.00	0.00	0.13	0.18	5
$d_x = 0.60$	R	1.00	1.00	0.00	0.00	100
$d_y = 0.60$	MR	0.02	0.02	0.14	0.19	7
$d_x = 0.85$	R	1.00	1.00	0.00	0.00	100
$d_y = 0.85$	MR	0.03	0.03	0.14	0.19	9
$d_x = 1.00$	R	1.00	1.00	0.00	0.00	100
$d_y = 1.00$	MR	0.00	0.00	0.13	0.18	5

Table:7-9 summarize distribution of R and MR for different levels of mutual correlations. It should be noted that powers of R and MR can not be compared since size of two is not same. However it is obvious from the Tables that MR is sharp enough to find out correlation between two series even with week correlation present between two series.

8 Conclusion

Correlation coefficient is a popular statistics in classical econometrics since it provides quick idea of the strength of association among the two variables. However it is now well recognized fact that the correlation coefficient R is strongly biased toward finding high correlation among the time series even if the two series are mutually uncorrelated. This

Table 4: : Distribution of R and MR for Different values of autoregressive parameter and different type of stationary: DGP: M1

		Mean	Median	SD	IQR	RR
$d_x = 1.00$	R	0.00	0.00	0.14	0.19	10
$d_y = 0.00$	MR	0.00	0.00	0.16	0.23	12
$d_x = 0.60$	R	0.00	0.00	0.23	0.32	32
$d_y = 0.85$	MR	0.00	0.00	0.17	0.24	13

Table 5: Distribution of R and MR for Different values of autoregressive parameter and different type of stationary:DGP: M2

		Mean	Median	SD	IQR	RR
$d_x = 1.00$	R	0.02	0.02	0.14	0.19	10
$d_y = 0.00$	MR	0.00	0.00	0.17	0.23	13
$d_x = 0.60$	R	0.06	0.07	0.23	0.32	35
$d_y = 0.85$	MR	0.00	0.00	0.17	0.24	14

Table 6: Distribution of R and MR for Different values of autoregressive parameter and different type of stationary:DGP: M3

		Mean	Median	SD	IQR	RR
$d_x = 1.00$	R	0.96	0.96	0.01	0.01	100
$d_y = 0.00$	MR	0.00	0.00	0.16	0.22	11
$d_x = 0.60$	R	1.00	1.00	0.00	0.00	100
$d_y = 0.85$	MR	0.03	0.02	0.17	0.24	15

Table 7: Distribution of R and MR for various degrees of mutual and autocorrelation

		No Correlation			Weak Correlation			Strong Correlation		
		M3			M6A			M6B		
		Mean	SD	OUT	Mean	SD	OUT	Mean	SD	OUT
$d_x = 0.00$	R	0.00	0.12	5	0.24	0.11	52	0.69	0.07	100
$d_y = 0.00$	MR	0.00	0.13	5	0.23	0.12	44	0.69	0.07	100
$d_x = 0.60$	R	0.00	0.17	17	0.24	0.16	52	0.69	0.09	100
$d_y = 0.60$	MR	0.00	0.13	5	0.23	0.13	45	0.69	0.07	100
$d_x = 0.85$	R	0.00	0.27	41	0.23	0.25	56	0.68	0.15	99
$d_y = 0.85$	MR	0.00	0.14	7	0.23	0.13	45	0.68	0.08	100
$d_x = 1.00$	R	0.00	0.49	72	0.22	0.47	75	0.66	0.31	91
$d_x = 1.00$	MR	0.00	0.15	8	0.23	0.14	43	0.68	0.09	100

Table 8: Distribution of R and MR for various degrees of mutual and autocorrelation

		No Correlation			Weak Correlation			Strong Correlation		
		M3			M6A			M6B		
		Mean	SD	OUT	Mean	SD	OUT	Mean	SD	OUT
$d_x = 0.00$	R	0.00	0.12	5	0.24	0.11	53	0.69	0.07	100
$d_y = 0.00$	MR	0.00	0.13	5	0.23	0.12	45	0.69	0.07	100
$d_x = 0.60$	R	0.02	0.17	17	0.25	0.16	56	0.69	0.09	100
$d_y = 0.60$	MR	0.00	0.13	5	0.23	0.13	44	0.69	0.07	100
$d_x = 0.85$	R	0.14	0.26	46	0.36	0.23	73	0.74	0.13	100
$d_y = 0.85$	MR	0.00	0.14	7	0.23	0.13	45	0.69	0.08	100
$d_x = 1.00$	R	0.94	0.05	100	0.96	0.03	100	0.99	0.02	100
$d_y = 1.00$	MR	0.01	0.15	8	0.23	0.14	45	0.68	0.09	100

Table 9: Distribution of Various Measures of Correlation for various degrees of mutual and autocorrelation

		No Correlation			Weak Correlation			Strong Correlation		
		M3			M6A			M6B		
		Mean	SD	OUT	Mean	SD	OUT	Mean	SD	OUT
$d_x = 0.00$	R	0.99	0.00	100	0.99	0.00	100	1.00	0.00	100
$d_y = 0.00$	MR	0.00	0.13	5	0.24	0.13	45	0.69	0.07	100
$d_x = 0.60$	R	1.00	0.00	100	1.00	0.00	100	1.00	0.00	100
$d_y = 0.06$	MR	0.02	0.14	7	0.24	0.13	48	0.68	0.08	100
$d_x = 0.85$	R	1.00	0.00	100	1.00	0.00	100	1.00	0.00	100
$d_y = 0.85$	MR	0.03	0.15	9	0.26	0.14	52	0.70	0.08	100
$d_x = 1.00$	R	1.00	0.00	100	1.00	0.00	100	1.00	0.00	100
$d_y = 1.00$	MR	0.00	0.13	5	0.24	0.12	47	0.70	0.07	100

phenomenon is usually called spurious correlation. We have presented the evidences that spurious correlation exist even among the stationary series. This means for time series models, in no case correlation provides reliable measure of association because almost all of economic time series are serially dependent. We developed a new measure of association MR which is robust to type and strength of autocorrelation, type of stationarity and type of deterministic part in the data generating process of the two series. MR is not formal test of association between two series rather it is a descriptive giving the immediate idea of the strength of association between two series. We recommend use of MR in the time series models instead of the conventional R.

9 Application

The modified R is assumed to give the quick idea of relationship between two variables. It should have ability to discriminate between spurious time series relationship and the genuine relationship. Usually its not possible to evaluate to evaluate any statistical procedure using real data because we do not know the nature of true relationship between real time series. However, the data on consumption and income provides opportunity for evaluating measures of correlation. If we take income and consumption of same country, they should have genuine correlation, but if we have consumption of one country versus income of another country, the correlation should not be strong. If we see that consumption of one country has very high correlation with income of some other country, this correlation is spurious. Table 10 summarizes computed R and MR between income and consumption of 15 OECD countries. The diagonal entries of Table 10 correspond to computed R and MR statistics for income and consumption of same country. We observe that the value of MR for almost all the diagonal entries is greater than 30 percent, showing that both R and MR suggest correlation between consumption. The off-diagonal entries of Table 10 correspond to calculated R and MR for different countries which should not have genuine correlation. Most of the off diagonal entries of MR statistics are smaller than 30 percent in magnitude indicating that there is no correlation between two series. On the other hand the R statistics is above 95 percent even for the non-

Table 10: Correlation between income and consumption of various countries

		Australia	Austria	Belgium	Denmark	Finland	France	Greece	Ireland	Italy
Australia	MR	**0.51**	-0.26	0.1	0.22	0.18	0	0.11	0.31	0.01
	R	**1**	0.99	0.98	0.98	0.98	0.99	0.98	0.99	0.98
Austria	MR	0.11	**0.59**	0.31	0.15	0.1	-0.18	0.28	0.34	0.22
	R	1	**1**	1	0.99	0.99	1	0.99	0.98	0.99
Belgium	MR	0.23	0.08	**0.36**	0.2	0.42	0.19	0.3	0.32	0.53
	R	1	1	**1**	0.99	0.99	1	0.99	0.98	0.99
Denmark	MR	0.24	0.02	0.28	**0.72**	0.23	0.23	0.14	0.19	0.19
	R	0.99	0.99	0.99	**0.99**	0.98	0.99	0.98	0.98	0.98
Finland	MR	0.41	-0.18	0.21	0.15	**0.73**	0.17	0.16	0.23	0.36
	R	0.99	0.99	0.99	0.98	**1**	0.99	0.99	0.98	0.99
France	MR	0.2	0.01	0.41	0.2	0.35	**0.22**	0.5	0.26	0.53
	R	0.99	1	1	0.99	0.99	**1**	1	0.97	1
Greece	MR	0.17	-0.04	0.39	0.28	0.26	0.21	**0.6**	0.51	0.04
	R	0.97	0.97	0.99	0.99	0.98	0.98	**0.99**	0.95	0.98
Ireland	MR	0.25	-0.15	0.14	-0.06	0.1	0.03	0.39	**0.45**	-0.01
	R	0.98	0.96	0.95	0.95	0.95	0.96	0.94	**0.99**	0.94
Italy	MR	0.24	0.04	0.58	0.27	0.37	0.19	0.21	0.22	**0.61**
	R	0.99	1	1	0.99	1	1	1	0.97	**1**

diagonal entries. This shows that R statistics over estimates strength of relationship between two series.

10 Directions for Further Research

Classical regression techniques when applied to time series data often produce spurious results. The conventional measures of relationship between variables e.g. t-statistics and R-square depends on IID assumption and this assumption is often invalid for time series. It has been proven that if two series contain unit root, the t-statistics and R-square are biased toward non-rejection of relationship between time series and this bias increases asymptotically (Granger and Newbold, 1974; Phillips, 1987). This problem creates difficulty to differentiate between correctly specified genuine economic relationship and the spurious relationship. Correlation coefficient is a popular statistics in classical econometrics since it provides quick idea of the strength of association among the two variables. We have presented the evidences that spurious correlation exist even among the stationary series. This means for time series models, in no case correlation provides reliable measure of association because almost all of economic time series are serially dependent. We developed a new measure of association MR which is robust to type and strength of autocorrelation, type of stationarity and type of deterministic part in the data generating process of the two series.

Acknowledgment

Author's thankful to Dr. Asad Zaman, DG IIIE, International Islamic University, Islamabad for his continuous guidance on earlier drafts of this paper.

References

Aldrich, J. (1995). Correlations genuine and spurious in pearson and yule. *Statistical Science*, pages 364–376.

Elderton, W. P. (2011). *Frequency curves and correlation*. Cambridge University Press.

Enders, W. (2008). *Applied econometric time series*. John Wiley & Sons.

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

Granger IV, C. W., Hyung, N., and Jeon, Y. (2001). Spurious regressions with stationary series. *Applied Economics*, 33(7):899–904.

Greene, W. H. (2003). *Econometric Analysis, 5/e*. Pearson Education India.

Gujarati, D. N. and Porter, D. C. (1999). Essentials of econometrics.

Nelson, C. R. and Plosser, C. R. (1982). Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of monetary economics*, 10(2):139–162.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110.

Pearson, K., Lee, A., and Bramley-Moore, L. (1899). Mathematical contributions to the theory of evolution. vi. genetic (reproductive) selection: Inheritance of fertility in man, and of fecundity in thoroughbred racehorses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 192:257–330.

Phillips, P. C. (1987). Time series regression with a unit root. *Econometrica: Journal of the Econometric Society*, pages 277–301.

Simon, H. A. (1954). Spurious correlation: A causal interpretation*. *Journal of the American Statistical Association*, 49(267):467–479.

Yule, G. U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society*, 60(4):812–854.

Yule, G. U. (1910). On the distribution of deaths with age when the causes of death act cumulatively, and similar frequency distributions. *Journal of the Royal Statistical Society*, 73(1):26–38.

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series?–a study in sampling and the nature of time-series. *Journal of the royal statistical society*, 89(1):1–63.