Using the Durbin-Watson Ratio to Detect a Spurious Regressions: Can We Make a Rule of Thumb?

M. Fukushige[a]

[a] *Graduate School of Economics, Kobe University, Nada-ku, Kobe, 657-8501, Japan* (mfuku@kobe-u.ac.jp)

Abstract: Recent advances in time series analysis related to unit root and cointegration make us nervous about spurious regressions. Spurious regression refers to the case where some statistically significant coefficients are often obtained in regression analysis when the dependent and independent variables are mutually independent random walks. High R-squared and significant t-values might mislead us to nonsense regressions. In this paper, I focus on the Durbin-Watson (DW) ratio and investigate whether it is helpful in detecting spurious regressions in empirical analyses or not. This statistic has a useful property to discriminate cointegrating relationships and spurious regression. In the case of a spurious regression, the DW ratio converges to zero, so it can be utilized for testing the hypothesis of a cointegrating relationship. However, there are several types of nonsense regressions. In some cases, the asymptotic distributions of the t-value, DW ratio or other statistics have been studied, but how the statistics might be utilized in testing or detecting a nonsense regression has not been investigated. In order to investigate if a traditional diagnostic testing process using t-values and the DW ratio works, I conduct a Monte Carlo simulation. The results show that the traditional diagnostic testing process when we check t-values and the DW ratio in the first regression and then proceed to cope with the serial correlation in disturbances works, and we can detect nonsense regressions when the spurious effect arising from non-stochastic part is removed. If the spurious effect is not removed, we have a statistically significant coefficient even in the second regression (Cochrane=Orcutt method). This result suggests that we should pay attention to whether the sources of spurious effects can be removed in the second regression or not. In other words, to apply a proper model to data is one of the most important step in detecting a spurious regression.

Keywords: Spurious regression, Unit root, Durbin-Watson ratio

1. INTRODUCTION

From an empirical researcher's viewpoint, recent advances in time series analysis related to unit root and cointegration make us nervous about spurious regression [Granger and Newbold, 1974]. In the case of a spurious regression, some statistically significant coefficients are obtained and the R-square is very high. This high R-square and significant t-values might mislead us to nonsense regressions. Only the Durbin-Watson (DW) ratio is a clue to detect a nonsense regression because its value is low. In this paper, I focus on this DW ratio and investigate whether it is helpful in detecting drop spurious regressions in empirical analyses or not.

The DW ratio is a one of the most popular statistics for testing for serial correlations in regression disturbances. This statistic also has a useful property to discriminate cointegrating relationships and spurious regressions. In the case of spurious regression, the DW ratio converges to zero [Phillips, 1986]. Engle and Granger [1987] proposed to utilize this property for testing the hypothesis of cointegrating relationship. Additionally, some other researchers investigate the properties of the DW ratio in some similar situations, for example, Bhargava [1986] and Hisamatsu and Maekawa [1994].

Of course, there are several types of nonsense regression apart from the case Granger and Newbold [1974] considered. In some cases, the asymptotic distributions of t-value, the DW ratio or other statistic have been studied, but how to utilize the statistics for testing or detecting a nonsense regression has not been investigated. In this paper, a traditional diagnostic testing process using t-values and the DW ratio is investigated. Conducting a Monte Carlo simulation for some nonsense regression cases, I focus on whether the traditional diagnostic testing process can be used as a rule of thumb or not.

2. SPURIOUS EFFECTS IN NONSENSE REGRESSIONS

Granger and Newbold [1974] consider the estimation of a simple regression:

$$\text{Model 1: } Y_t=c_0+c_1X_t+e_t \quad (1)$$

where $t=1,\ldots,T$, and Y_t and X_t are assumed to be generated as two independent random walks:

$$Y_t=Y_{t-1}+u_t \text{ and } X_t=X_{t-1}+v_t. \quad (2)$$

They found that the calculated t-values for c_1 in (1), t_{c1}, are significant with relatively high frequencies, and that the DW ratios are low. This regression is called a "spurious regression."

Phillips [1986] investigated the asymptotic distribution of the DW and t_{c1} and obtained the following results:

$$DW = O_p(T^{-1}),$$

$$t_{c1} = O_p(T).$$

These results means that the calculated t-values diverge, while the DW ratio converges to zero.

This type nonsense regression has been extended to several more complicated cases; foe example, to the case with cointegrated regressors in Choi [1994], and panel data in Entorf [1997] and Kao [1999].

In addition to data generation process in (1), there are several types of data generation processes that make estimation of a regression models nonsensial. Consider the case where Y_t and X_t follow two independent random variables without or with trend:

$$Y_t=B1+u_t \text{ and } X_t=B2+v_t \quad (3)$$

$$Y_t=B1\,t+u_t \text{ and } X_t=B2\,t+v_t, \quad (3')$$

where B1 and B2 are parameters, and t is a trend variable. We can take a regression like (1) as a nonsense regression because we cannot discriminate the theory-based (*e.g.* economic theory) relationships from a just computational or statistical relationship. Of course, if some shifts or kinks occurred in the value of the parameters with time leads or lags, we can define causality between the two variables and a regression like (1) has some meaning.

Moreover, we can consider more complicated cases. Suppose Y_t and X_t follow two independent random walks with drift:

$$Y_t=B1+Y_{t-1}+u_t \text{ and } X_t=B2+X_{t-1}+v_t. \quad (4)$$

For this case, we should take the following model:

$$\text{Model 2 } Y_t=c_0+c_1X_t+c_2\,t+e_t \quad (5)$$

as a kind of nonsense regression. Entorf [1987] investigated the properties of statistics for Model 1 with the data generating process (4). He obtained the following results:

$$DW = O_p(T^{-1}),$$

$$t_{c1} = O_p(T),$$

This is the same result as obtained by Phillips [1986] for the Granger-Newbold model.

Some of the asymptotic results of these nonsense regressions are intuitive and others are obtained theoretically. However, their small sample properties are obscure, and the properties of the traditional diagnostic testing process have not been investigated.

3. DW AND t-RATIO IN PRACTICE

Here, I focus on the roles of the DW ratio in traditional diagnostic testing process. What we call "traditional diagnostic testing process" is as follows:

1st: Test the null hypothesis that the regression coefficients are not significant using t-statistics.

2nd: Test the null hypothesis that there is no serial correlation in the error terms using the DW ratio.

3rd: When the null hypothesis in step 2 is rejected, we proceed to cope with the serially correlated error terms by applying Cochrane=Orcutt procedure to the data. Then, we test the null hypothesis that the coefficients are not significant using t-statistics.

If this diagnostic testing process works to drop the nonsense regression, we need not pay attention to whether the time series has unit root or not preliminary. We should just test the cointegrating relationships with Johansen's [1988] method.

4. MONTE CARLO SIMULATION

First, we need to define the data generating processes (DGP). Because the results with DGP (3) and (3') are intuitive, we adopt (2) and (4) as the DGP for simulation. Following Entorf [1992] that is cited in Maddala and Kim [p.181, 1998], I fix the ratio B1/B2=2.0 except for DGP:X4-, and in all the cases, the error variance of u_t and v_t are fixed 1.0. Details of the DGPs (sample size, *etc*) are shown in Table 1.

Table 1. Setups of the models.

DGP	Number of observations	B1	B2
X0-50	50	0	0
X0-100	100	0	0
X0-150	150	0	0
X0-200	200	0	0
X1-50	50	0.1	0.05
X1-100	100	0.1	0.05
X1-150	150	0.1	0.05
X1-200	200	0.1	0.05
X2-50	50	0.5	0.25
X2-100	100	0.5	0.25
X2-150	150	0.5	0.25
X2-200	200	0.5	0.25
X3-50	50	1.5	0.75
X3-100	100	1.5	0.75
X3-150	150	1.5	0.75
X3-200	200	1.5	0.75
X4-50	50	0.5	0
X4-100	100	0.5	0
X4-150	150	0.5	0
X4-200	200	0.5	0

There are several variations of which significance levels and which statistics d_L or d_U for the DW we should utilize. In this paper, I adopt the following steps:

1st: I reject the null hypothesis that the coefficient c_1 in (1) or (5) equal zero if its corresponding t-statistic is larger than 1.96 in absolute value, which is the 5% critical value from standard normal distribution and then proceed to the second step.

2nd: I reject the null hypothesis that there is no serial correlation in the error term when the calculated the DW is less than d_L, then proceed to the third step: the Cochrane=Orcutt procedure.

3rd: If the result from the Cochrane=Orcutt procedure show that the t-value of the coefficient c_1 in (1) or (5) is larger than 1.96 in absolute value, I reject the null hypothesis that the coefficient is zero. Then, I conclude that this regression is not a nonsense regression.

I set the number of replications at 5000. The results of applying this diagnostic testing process to Model 1 (equation (1)) are shown in Table 2. t1, t1 & DW, and t1, DW & and t2 columns show the percentages of rejection rate in the first, second and third steps, respectively.

The results can be summarized as follows: First, the rejection rates in the first and second steps are almost the same and more than 60%. Second, except in the X2- and X3- cases, the rejection rates in the third step are around 5%. This is close to the target-size of the testing in the third step. In the X2- and X3- cases, the rejection rates are high. Especially, in the X3- cases, because of their high rejection rates in the third steps, we can conclude that this diagnostic testing process cannot detect spurious regressions when we estimate the Model 1 with DGP (4), random walk with drift. The reason why this process produces misleading results depends on the spurious effect from the nonstochastic part (drift) in (4). Consider the case of applying the diagnostic testing process for (3) or (3'). Because of the nonstochastic part, estimated coefficients become statistically significant. To avoid this type misleading, we should add a nonstochastic trend in the estimated model.

Table 2. Simulation results: Model 1.

DGP	t1	t1 & DW	t1, DW & t2
X0-50	66.0	66.0	8.6
X0-100	75.6	75.6	5.4
X0-150	81.8	81.8	4.9
X0-200	83.3	83.3	5.0
X1-50	67.8	67.8	6.9
X1-100	81.0	81.0	6.4
X1-150	86.5	86.5	5.0
X1-200	88.1	88.1	5.7
X2-50	94.3	94.3	7.9
X2-100	99.1	99.1	7.6
X2-150	99.6	99.6	7.6
X2-200	99.7	99.7	8.7
X3-50	100.0	100.0	26.8
X3-100	100.0	100.0	52.4
X3-150	100.0	100.0	71.9
X3-200	100.0	100.0	74.9
X4-50	80.9	80.9	4.3
X4-100	85.8	85.8	5.0
X4-150	90.6	90.6	4.0
X4-200	93.1	93.1	5.2

Therefore, I conduct a simulation of the properties of Model 2 (equation (5)). Because a nonstochastic trend is added, the sources of spurious effects are removed: a stochastic trend is detected by the DW ratio and the nonstochastic trend removed by trend term. The results of the simulations are shown in Table 3.

The results can be summarized as follows: First, the rejection rates in the first and second steps are almost the same and more than 60%. These are similar to the results in Table 2. Second, even in the X2- and X3- cases, the rejection rates in the third step are around 5%. This is close to the target-size of the testing procedure in the third step. This result means that the traditional diagnostic testing process works and can detect the spurious regressions at proper rates even when the random walks have drifts. It is clear that adding trend term removes the spurious effect from the nonstochastic part, and .the size of the third step is close to 5%.

Table 3. Simulation results: Model 2.

DGP	t1	t1 & DW	t1, DW & t2
X0-50	51.7	51.5	7.1
X0-100	64.7	64.7	5.0
X0-150	70.5	70.5	3.8
X0-200	70.8	70.8	4.7
X1-50	52.8	52.8	4.7
X1-100	75.6	75.6	5.0
X1-150	81.8	81.8	3.8
X1-200	83.3	83.3	4.7
X2-50	52.8	52.8	7.0
X2-100	64.7	64.7	5.0
X2-150	70.5	70.5	3.8
X2-200	70.8	70.8	4.7
X3-50	52.8	52.8	7.0
X3-100	64.7	64.7	5.0
X3-150	70.5	70.5	3.8
X3-200	70.8	70.8	4.7
X4-50	52.8	52.8	7.0
X4-100	64.7	64.7	5.4
X4-150	70.5	70.5	4.9
X4-200	70.8	70.8	5.0

Comparing the results from Table 2 and 3, one of the most important implications that we can detect a spurious regression when we deal with both sources of spurious effects: the stochastic and nonstochastic ones. In practice, we should add trend term to remove the nonstochastic one and check the DW ratio to detect the stochastic one.

5. CONCLUSION

I conduct a Monte Carlo simulation for some nonsense regressions. The results show that the traditional diagnostic checking process: t-values and the DW ratio in the first regression and t-value in the second regression, detect nonsense regression when the spurious effect from the non-stochastic part is removed. If a spurious effect is not removed, we have a statistically significant coefficient in the second regression (Cochrane=Orcutt method).

This result suggests that we should pay attention to whether the sources of spurious effects can be removed in the second regression or not. In other words, to apply a proper model to data is one of the most important to detect a spurious regression.

Maddala and Kim [1998] introduce Entorf [1992] results as:

> "Entorf investigates the small sample distribution of the DW statistic. He finds that for more than five regressions the 95 percent fractile of the DW distribution is larger than two. Thus even regressions with the DW values of about two do not necessarily ensure that we do not estimate spurious regressions."

Because I do not obtain Entorf's [1992] paper, I do know the exact setup of his simulation experiments. However, we should investigate the case with multiple explanatory variables. Of course, Granger and Newbold [1974] investigated such cases, and their result shows that the DW ratio becomes large in accordance with increase of explanatory variables.

6. REFERENCES

Bhargava, A., On the theory of testing for unit roots in observed time series, *Review of Economic Studies*, 53, 369-384, 1986.

Choi, I., Spurious regressions and residual based tests for cointegration when regressors are cointegrated, *Journal of Econometrics*, 60, 313-320, 1994.

Engle, R., and C.W.J. Granger, Co-integration and error correction: representation, estimation and testing, *Econometrica*, 55, 251-276, 1987.

Entorf, H., Random walks with drifts, simultaneous equation errors, and small sample properties: simulating bird-eye view, INSEE Discussion Paper, No.9222, 1992.

Entorf, H., Random walks with drifts: nonsense regression and spurious fixed-effect estimation, *Journal of Econometrics*, 80, 287-296, 1997.

Granger, C.W.J., and P. Newbold, Spurious regressions in econometrics, *Journal of Econometrics*, 2, 111-120, 1974.

Hisamatsu, H., and K. Maekawa, The distribution of the Durbin-Watson statistic in integrated and near-integrated models, *Journal of Econometrics*, 61, 367-382, 1994.

Johansen, S., Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, 12, 231-254, 1988.

Kao, C., Spurious regression and residual-based tests for cointegration in panel data, *Journal of Econometrics*, 90, 1-44, 1999.

Maddala, G.S., and I.-M. Kim, *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, 1998.

Phillips, P.C.B., Understanding spurious regressions in econometrics, *Journal of Econometrics*, 33, 311-340, 1986.