EVIDENCE OF ARCH(1) ERRORS IN THE CONTEXT OF SPURIOUS REGRESSIONS

Christos Agiakloglou

HAL Id: hal-00520668

https://hal.science/hal-00520668

Submitted on 24 Sep 2010

EVIDENCE OF ARCH(1) ERRORS IN THE CONTEXT OF SPURIOUS REGRESSIONS

EVIDENCE OF ARCH(1) ERRORS IN THE CONTEXT OF
SPURIOUS REGRESSIONS

Christos Agiakloglou

Department of Economics, University of Piraeus, Greece

Abstract

The spurious regression phenomenon is related to first-order serially correlated errors. This study, using a Monte Carlo analysis, finds that this phenomenon is also related to ARCH(1) type errors.

Keywords: Spurious regression, stationary and non-stationary processes, serially
          correlated errors, ARCH(1) errors

JEL Classification: C22

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

1. Introduction

When time series data is used to analyze economic behaviors it is very likely that serially correlated errors will appear in regression analysis. If this type of error structure is identified, it must be corrected, simply because serious statistical problems will arise in the regression analysis, if it is ignored. Thus, the typical question asked in this case is whether or not this problem is indeed a problem of autocorrelated errors so that it can be cured by using any standard procedure, such as the Cochrane–Orcutt. Otherwise, the analyst must search for other sources causing the presence of such behavior in the errors.

In the literature, the presence of autocorrelated errors is highly related to the spurious regressions phenomenon, introduced by Granger and Newbold (1974) and well-documented mathematically by Phillips (1986). By spurious regressions we refer to the issue of getting evidence of a linear relationship between two variables when they are not linearly related. Granger and Newbold (1974), using a Monte Carlo analysis, examined the behavior of this phenomenon and they found that when ordinary least squares is applied to two independent time series generated by integrated processes it is very likely that an analyst will get misleading statistical results, a phenomenon that it is also investigated by Granger *et al* (2001) for two independent stationary AR(1) processes. Moreover, Granger and Newbold (1974) pointed out that a low value of a Durbin-Watson statistic will appear in regression analysis as an indication of a false relationship between the two variables and standard procedures, such as the Cochrane-Orcutt, will fail to correct this problem. The presence of serially correlated errors, in the context of spurious regressions, is

also investigated by Newbold and Davies (1978) using a multivariate model in which

all variables were generated by non-stationary moving average processes.

The problem of serially correlated errors is also related to the Autoregressive

Conditionally Heteroscedastic (ARCH) models, presented by Engle (1982). In this

case, the problem of heteroscedasticy, in which the variance of the error term is

determined by an autoregressive scheme, will appear in regression analysis as a

problem of autocorrelation.

This study, using a Monte Carlo analysis, finds that the spurious regression

phenomenon is also related to ARCH(1) errors. In fact, strong evidence of ARCH(1)

errors will appear not only in the case of two independent variables, generated by

stationary or non-stationary processes, but also in the multivariate case. Therefore,

regression analysis that involves time series should carefully examine the issue of

serially correlated errors since autocorrelation can appear not only in the context of

first moments but also in the context of second moments and thus the classical

assumptions of constant variance and independent errors are both violated.

2. Simulation results for the bivariate case

To investigate whether or not a linear relationship exists between any two variables an

analyst will estimate, using OLS, the following equation:

$$Y_t = \alpha + \beta X_t + \varepsilon_t \tag{1}$$

where the error term is assumed to be *iid* normally distributed with mean zero and

constant variance σ_ε^2, i.e., $\varepsilon_t \sim iid\mathrm{N}(0, \sigma_\varepsilon^2)$. If there is a linear relationship between

these two variables, estimation of equation (1) should produce "good" statistical

results and the null hypothesis of β being zero should be rejected. However, as Granger and Newbold (1974) have pointed out, estimation of equation (1) using two independent random walk processes without drift will produce evidence of serially correlated errors and the null hypothesis of β being zero will be rejected surprising very frequently for moderate sample sizes. Thus, the classical assumption of independent error structure will be violated.

To illustrate, consider two independent processes Y_t and X_t generated from the following DGP:

$$Y_t = \varphi_y Y_{t-1} + \varepsilon_{yt} \tag{2}$$

$$X_t = \varphi_x X_{t-1} + \varepsilon_{xt} \tag{3}$$

where the errors ε_{yt} and ε_{xt} are white noise processes independent of each other and the autoregressive parameter is allowed to take values of 0.0, 0.2, 0.5, 0.8, 0.9 and 1.0. For all values of the autoregressive parameters that are less than one both processes are stationary, i.e., first-order autoregressive processes, whereas for $\varphi_y = \varphi_x = 1$ both processes are non-stationary, i.e., random walk processes without drift. In the case of $\varphi_y = \varphi_x = 0$ both processes are white noise.

Table 1 reports the proportions of rejections of the null hypothesis that $\beta = 0$ at the 5% nominal level, using the exact critical values of the t distribution, and Table 2 reports mean values of the Durbin–Watson statistic obtained from the residuals of the estimation of model (1) for series generated from equations (2) and (3) for sample sizes of 50 and 100 observations using 1,000 replications. As can be seen from these tables, the problem of serially correlated errors will appear in regression analysis along with spurious results including the case of stationary processes. However, autocorrelation does not appear only where the null hypothesis of $\beta = 0$ is frequently rejected. It appears also in cases where the null hypothesis is accepted. For example,

the analyst will get the same proportions of rejections of the null hypothesis that $\beta = 0$ regardless if he is regressing a random walk process against a white noise process or a white noise process against a random walk process, as Table 1 reports. In fact, spurious behavior does not depend on the type of the dependent variable used in regression and it treats both variables symmetrically. However, in the first case, the mean value of the Durbin-Watson statistic is close to zero, whereas in the second case the mean value of this statistic is close to two. Hence, strong evidence of serially correlated errors will appear only in the case of regressing a random walk process against a white noise process, as Table 2 reports. That means that the degree of evidence of autocorrelated errors depends on the true generating behavior of the dependent variable and not on the independent variable used in regression analysis, a statement that does not exist in the case of spurious regressions.

In addition, evidence of ARCH(1) errors will also appear in regression analysis. To investigate this behavior using the residuals of this simulation process the following equation is estimated:

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + u_t \tag{4}$$

where the error term u satisfies all standard regression assumptions. The test of ARCH(1) errors is based on the value of α_1 coefficient. If $\alpha_1 \neq 0$, the errors are generated by an ARCH(1) model, whereas if $\alpha_1 = 0$, the variance of the errors will be considered as constant. The test of the null hypothesis that $\alpha_1 = 0$ is curried out as a Lagrange Multiplier test, i.e., $LM = nR^2$, where R^2 is the coefficient of determination of equation (4) and n is the sample size.[1] The LM statistic is compared with the critical value obtained from the X^2 distribution with one degree of freedom and the

---

[1] For our simulation process n is considered as of 49 and 99 observations.

proportions of rejections of the null hypothesis of an ARCH(1) model at the 5%
nominal level are reported on Table 3.

Perhaps, the most astonishing feature of Table 3 is that evidence of ARCH(1)
errors will also appear in regression analysis in the context of spurious regressions
and in the same way as autocorrelation. Indeed, the null hypothesis that $\alpha_1 = 0$ will be
rejected with certainty, i.e., 100%, when the independent variable is a random walk
process even for moderate sample sizes of 100 observations. Similar behavior exists
and for stationary processes. For example, for two independent AR(1) processes with
an autoregressive parameter of 0.8 with 100 observations, the null hypothesis of $\beta = 0$
will be rejected 33.6%, the mean value of the Durbin-Watson statistic will be very
close to zero, i.e., 0.4966, indicating strong evidence of first-order serially correlated
errors and the null hypothesis of ARCH(1) errors will be accepted almost with
certainty, i.e., at 99.1% level. As a general remark, the degree of evidence of
ARCH(1) errors depends, as in the case of autocorrelated errors, on the structure pf
the dependent variable.

3. Simulation results for the multivariate case

To determine whether k independent variables explain the behavior of the dependent
variable Y_t an analyst must estimate, using OLS, the following regression model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + ... + \beta_k X_{kt} + \varepsilon_t \tag{5}$$

where the error term is assumed to be *iid* normally distributed with mean zero and
constant variance σ_ε^2, i.e., $\varepsilon_t \quad iidN(0, \sigma_\varepsilon^2)$, and it is not correlated with the
independent variables. The test to investigate whether the k independent variables

simultaneously explain the behavior of the dependent variable is based on the following null hypothesis:

$$H_0: \beta_1 = \beta_2 = ... = \beta_k = 0 \tag{6}$$

i.e., that all coefficients of model (5) are zero, against the alternative that at least one coefficient is not zero and it implemented using the conventional F statistic or the t statistic for the special case of $k = 1$.

This phenomenon is initially investigated by Newbold and Davies (1978) for non-stationary moving average processes and for values of $k = 1$ and 4 for sample sizes of 50 observations. Newbold and Davies (1978) showed that inference in regressions involving time series can be strongly affected by the error structure. In this paper the spurious regression phenomenon is investigated along with the error structure for stationary as well as for non-stationary processes as the number of independent variables increases from one to four.[2]

Thus, model (5) is estimated using one through four independent variables generated by the same DGP as in the bivariate case, for sample sizes of 50 and 100 observations and the null hypothesis that all coefficients of model (5) are zero, given by equation (6), is tested using the usual t statistic for $k = 1$ and the conventional F statistic for values of $k = 2$, 3, and 4. The results of this simulation process based on 1,000 replications are reported on Table 4.

Perhaps, the most interesting feature of Table 4 is the fact that for all values of the autoregressive parameter greater than zero the proportion of rejections of the null hypothesis that all coefficients of model (5) are zero increases as more independent variables are added to the regression analysis. This result suggests that the analyst will get more spurious results as k increases, regardless of the structure of the

---

[2] Note that for $k = 1$ the simulation process is already examined.

processes used in the regression analysis. For the random walk case, for example, the null hypothesis (6) is rejected 98.8% for $k = 4$ instead of 77.4% for $k = 1$ at the 5% nominal level for sample sizes of 100 observations, whereas using stationary AR(1) processes with $\varphi = 0.8$ these numbers are 68.8% and 33.6% respectively. Only for the white noise case the analyst will not find spurious results.

The next step is to investigate the error structure of this simulation process based on the residuals $\hat{\varepsilon}_t$ obtained from the estimation of model (5) for all values of k. For this purpose, the Durbin–Watson statistic is calculated to examine the presence of first–order serially correlated errors and the mean value of this statistic using 1,000 replications are reported on Table 5. Furthermore, the residuals of this simulation process are also used to calculate the LM statistic for testing for ARCH(1) errors and the proportions of rejections of the null hypothesis that $\alpha_1 = 0$ are reported on Table 6. As can be seen from these tables, strong evidence of autocorrelated errors in the first and in the second moments will also appear in regression analysis for stationary AR(1) process with moderate and large values of the autoregressive parameter, as well as for non-stationary processes, regardless of the number of independent variables used in the regression analysis. Only in the white noise case the errors are not serially correlated, since the mean value of the Durbin–Watson statistic is 2 and the proportions of rejections of the null hypothesis are close to the nominal levels.

4. Conclusions

This study examined the spurious regression phenomenon for stationary and non-stationary independent processes for the bivariate and the multivariate case in the

context of serially correlated errors. Using a Monte Carlo analysis this study finds

that the concept of serially correlated errors is a broader concept than the spurious

regression phenomenon and autocorrelated errors in first and in second moments,

based on an ARCH(1) model, will appear in regression analysis that uses time series

data. Indeed, for moderate sample sizes the null hypothesis of an ARCH(1) error

structure will be rejected all the time. Moreover, simulation results indicate that an

analyst will get more spurious results as the number of independent variables

increases.

References

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.

Newbold, P. and Davies, N., 1978, "Error mis-specification and spurious regressions", *International Economic Review*, 19, 513-519.

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

Table 1
Proportions of rejections of the null hypothesis that $\beta = 0$ at the 5% nominal level based on 1,000 replications

φ_y	φ_x					
	1.0	0.9	0.8	0.5	0.2	0.0
1.0	0.652	0.557	0.437	0.238	0.110	0.049
	0.774	0.612	0.466	0.228	0.099	0.046
0.9	0.556	0.474	0.385	0.217	0.091	0.044
	0.591	0.476	0.422	0.221	0.106	0.061
0.8	0439	0.388	0.323	0.182	0.085	0.048
	0.485	0.402	0.336	0.190	0.087	0.048
0.5	0.204	0.235	0.181	0.120	0.073	0.053
	0.220	0.222	0.192	0.139	0.075	0.053
0.2	0.078	0.088	0.076	0.086	0.053	0.054
	0.106	0.109	0.116	0.069	0.067	0.049
0.0	0.049	0.046	0.050	0.053	0.053	0.056
	0.056	0.056	0.044	0.041	0.053	0.047

Note: In each cell numbers are corresponding for sample sizes of 50 and 100 observations respectively.

Table 2
Mean values of the Durbin–Watson Statistic obtained from estimation of model (1) based on 1,000 replications

φ_y	φ_x					
	1.0	0.9	0.8	0.5	0.2	0.0
1.0	0.3192	0.3054	0.2847	0.2605	0.2440	0.2360
	0.1721	0.1516	0.1439	0.1336	0.1269	0.1230
0.9	0.4570	0.4494	0.4331	0.4135	0.3970	0.3862
	0.3230	0.3179	0.3150	0.3044	0.2992	0.2953
0.8	0.6151	0.6098	0.6012	0.5837	0.5678	0.5579
	0.5074	0.4998	0.4966	0.4884	0.4833	0.4783
0.5	1.1351	1.1377	1.1313	1.1155	1.1044	1.0933
	1.0648	1.0645	1.0609	1.0570	1.0487	1.0449
0.2	1.6655	1.6597	1.6557	1.6503	1.6374	1.6306
	1.6460	1.6455	1.645	1.6375	1.6316	1.6268
0.0	2.0253	2.0230	2.0175	2.0093	1.9970	1.9890
	2.0118	2.0105	2.0062	2.0026	1.9955	1.9927

Note: In each cell numbers are corresponding for sample sizes of 50 and 100 observations respectively.

Table 3

Proportions of rejections of the null hypothesis that $\alpha_1 = 0$ at the 5% nominal level obtained from estimation of model (1) based on 1,000 replications

φ_y	φ_x					
	1.0	0.9	0.8	0.5	0.2	0.0
1.0	0.946	0.984	0.962	0.968	0.973	0.978
	1.000	1.000	1.000	1.000	1.000	1.000
0.9	0.883	0.868	0.893	0.896	0.911	0.919
	0.999	0.997	0.999	0.999	0.999	0.999
0.8	0.748	0.727	0.747	0.781	0.807	0.799
	0.982	0.986	0.991	0.991	0.990	0.990
0.5	0.201	0.197	0.199	0.211	0.220	0.233
	0.450	0.441	0.446	0.444	0.457	0.463
0.2	0.039	0.037	0.042	0.039	0.041	0.048
	0.064	0.075	0.069	0.070	0.071	0.072
0.0	0.031	0.035	0.034	0.032	0.034	0.030
	0.039	0.043	0.040	0.040	0.044	0.045

Note: In each cell numbers are corresponding for sample sizes of 50 and 100 observations respectively.

Table 4

Proportion of rejections of the null hypothesis that all coefficients of (5) are zero for values of k = 1, 2, 3 and 4 at the 5% nominal level based on 1,000 replications

φ	k			
	1	2	3	4
1.0	0.652	0.853	0.923	0.956
	0.774	0.917	0.965	0.988
0.9	0.474	0.669	0.781	0.855
	0.476	0.715	0.795	0.870
0.8	0.323	0.460	0.569	0.665
	0.336	0.523	0.606	0.688
0.5	0.120	0.155	0.182	0.192
	0.139	0.166	0.191	0.226
0.2	0.053	0.058	0.063	0.072
	0.067	0.057	0.054	0.071
0.0	0.056	0.063	0.059	0.052
	0.047	0.054	0.059	0.059

Note: In each cell numbers are corresponding for sample sizes of 50 and 100 observations respectively.

Table 5
Mean values of the Durbin–Watson Statistic obtained from estimation of model (5) based on 1,000 replications

φ	k			
	1	2	3	4
1.0	0.3192	0.4639	0.5895	0.7035
	0.1721	0.2444	0.3131	0.3848
0.9	0.4494	0.5444	0.6438	0.7436
	0.3179	0.3641	0.4114	0.4588
0.8	0.6012	0.6815	0.7603	0.8389
	0.4966	0.5358	0.5692	0.6047
0.5	1.1155	1.1585	1.2029	1.2415
	1.0570	1.0761	1.0970	1.1179
0.2	1.6374	1.6559	1.6713	1.6892
	1.6316	1.6393	1.6475	1.6569
0.0	1.9890	1.9870	1.9866	1.9881
	1.9927	1.9941	1.9944	1.9964

Note: In each cell numbers are corresponding for sample sizes of 50 and 100 observations respectively.

Table 6
Proportion of rejections of the null hypothesis that $\alpha_1 = 0$ at the 5% nominal level obtained from estimation of model (5) based on 1,000 replications

φ	k			
	1	2	3	4
1.0	0.946	0.857	0.747	0.628
	1.000	1.000	0.998	0.993
0.9	0.868	0.777	0.678	0.575
	0.997	0.995	0.991	0.981
0.8	0.747	0.668	0.578	0.471
	0.991	0.984	0.972	0.956
0.5	0.211	0.188	0.163	0.146
	0.444	0.430	0.394	0.383
0.2	0.041	0.039	0.037	0.036
	0.071	0.071	0.066	0.068
0.0	0.030	0.028	0.027	0.027
	0.045	0.049	0.047	0.048

Note: In each cell numbers are corresponding for sample sizes of 50 and 100 observations respectively.