The Use and Misuse of Summary Statistics in Regression Analysis

By Robert V. Bishop*

Abstract

This article discusses the effect of an autocorrelated error structure on the interpretation of traditional significance tests, especially the t-test and R^2 measure It emphasizes first-order serial correlation, a common and often serious problem that researchers using time series data may encounter Even though many of the problems associated with an autocorrelated error structure are well known, many researchers ignore them and report results which range from being potentially misleading to grossly erroneous

Keywords

Regression analysis, Autocorrelation, Filtering, Summary statistics

Introduction

In this article, I survey recent methodological developments concerning error structures which are "contaminated" with autocorrelation[1] and draw implications relevant for the interpretation and application of empirical econometric research

It is common to find instances where researchers simply report Durbin-Watson statistics[2] that suggest an error structure which is first-order autocorrelated[3] without taking account of this when interpretating their results A bias is introduced into the traditional tests of significance when autocorrelated errors are present,[4] this suggests that interpretation of these tests under autocorrelation is difficult, if not impossible Furthermore, recognizing and correcting for first-order serial correlation with the usual Cochrane-Orcutt or Hildreth-Lu procedure is inadequate in certain rather common situations (*5, 16*) [5]

If the error structure exhibits first-order autocorrelation and we assume that no relevant variable has been omitted,[6] the estimated regression coefficients are unbiased and consistent (as is well known), but they possess the undesirable property of being inefficient [7] What is also well documented in the literature, but often overlooked in practice, is that the usual tests of significance, when performed in the presence of autocorrelated errors, are biased For example, if positive first-order autocorrelation is present in the error structure and the independent variable is also autocorrelated, the estimates of the standard errors on each of the coefficients ($s_{\hat{\beta}}$) will be biased downward in most situations When the standard error of the coefficient is underestimated, the *t*-statistic on that coefficient is obviously overstated as it is computed as $t = \hat{\beta}/s_{\hat{\beta}}$, implying greater explanatory power for that variable

*The author was formerly an economist in the International Economics Division, ESS, and is currently an econometrician with Hershey Foods Corporation

[1] By definition, autocorrelation is serial correlation in time series data Serial correlation can also exist in cross-sectional data, such as spatial correlation (across geographic regions) or mutual correlation (across groups, such as income)

[2] The Durbin-Watson statistic (d) is computed as

$$d = \frac{\Sigma(u_t - u_{t-1})^2}{\Sigma u_t^2}$$

This approximates $d = 2(1 - \rho)$ where ρ is the estimated first order autocorrelation coefficient of the residuals in the model As $\rho \rightarrow 0$, $d \rightarrow 2$ By examining a table for upper (d_u) and lower (d_l) bounds on this statistic for the appropriate degrees of freedom, the researchers can test the null hypothesis of no autocorrelation

[3] First-order autocorrelation implies that the error term in period t is correlated with the error term in period t − 1, that is

$$e_t = \rho e_{t-1} + u_t$$

where $E(e_t u_t) = 0$

[4] The bias referred to here arises due to the consistent under- or over-estimation of the variances of the estimated coefficients The estimation of the variance-covariance matrix of the estimators (β's) is computed as $s_e^2\ (X'X)^{-1}$ which does not include the information embodied in the off-diagonal elements of the variance-covariance matrix of the disturbance terms commonly referred to as Ω in Generalized Least Squares applications The true variance-covariance matrix of the β's is given by $s_e^2\ (X'\Omega^{-1}X)^{-1}$

[5] Italicized numbers in parentheses refer to items in the references at the end of this article

[6] If a relevant causal variable has been omitted, the estimates are also biased

[7] The properties of the estimators in the presence of autocorrelation are discussed in many good econometrics texts, for example (*10*, pp 273-82)

than actually exists This situation can easily lead to the inclusion of a statistically irrelevant variable in the final model If the error structure exhibits negative serial correlation and the independent variable is positively autocorrelated, the standard errors of the coefficients are likely to be overestimated, possibly leading to the elimination of a statistically significant variable from the model [8]

Granger and Newbold performed a series of tests in which they examined the potential for discovering "spurious" relationships due to problems with autocorrelated errors(*12*) They believed that much econometric work documented in the literature was permeated with "relationships" which existed due only to the researcher's failure to remove autocorrelation from the error structure They examined the values of the coefficient of determination (R^2) generated by regressions in a Monte Carlo experiment Two independent series were generated, one a random walk and the other a more complicated autoregressive, integrated, moving-average structure, specifically ARIMA (0, 1, 1) structure [9] Granger and Newbold concluded that

> It is quite clear from these simulations that if one's variables are random walks and one includes in regression equations variables which should in fact not be included, then *it will be the rule* rather than the exception to find spurious relationships It is also clear that a high value for R^2 or $\bar{R}^2$ combined with a low value of *d* (Durbin-Watson statistic) is no indication of a true relationship (*12*)

In a later article, these authors further elaborated their position

> In time series regressions involving the levels of economic variables, one frequently sees coefficients of multiple correlation (R^2) much higher than 0 9 If these indicate anything at all, they presumably imply an extremely strong relationship between the dependent variable and the independent variables This is extremely misleading on many occasions, as comments noting poor forecast performance which sometimes follow these equations will testify In fact, the high R^2 values could be no more than a reflection of the fact that the dependent variable is highly autocorrelated and could easily be achieved simply by regressing the variable on its own past Thus, in such circumstances, the value of R^2 says nothing at all about the strength of the relationship between the dependent and independent variable (*13*)

If the error structure is first-order autoregressive (AR(1)),[10] the ordinary least squares (OLS) estimates of the regression parameters are (1) unbiased, (2) consistent, but (3) inefficient in small as well as in large samples The estimates of the standard errors of the coefficients in a model are biased downward if the residuals are positively autocorrelated and the independent variable itself is positively autocorrelated, they are biased upward if the residuals are negatively autocorrelated and the independent variable is positively autocorrelated Therefore, the calculated t-statistic is biased upward or downward in the opposite direction of the bias in the estimated standard error of that coefficient Granger and Newbold have demonstrated that the R^2 measure (both adjusted and unadjusted)[11] is usually grossly misleading in the presence of an autocorrelated error structure (*12*) They have further suggested that the regression results can be defined as "nonsense" if the R^2 measure exceeds that computed for the Durbin-Watson statistic (*13*)

To demonstrate how misleading regression statistics can be, I offer an example The natural logarithm of the quarterly measure of the U S consumer price index (CPI) was regressed on the logarithm of the narrowly defined U S money stock (old M1) for the period 1947-78 This is a test of a model describing the Crude Quantity Theory of Money, which states that changes in the exogenous money stock cause changes in the passive (endogenous) price level

[8] One should not view negative autocorrelation as a mirror-image of positive autocorrelation The different results obtained under positive and negative autocorrelation are due to the direction of the bias in the estimate of the standard error of the coefficients under the two conditions

[9] A process characterized as ARIMA (0, 1, 1) is an integrated (regular differencing for stationarity is applied), mixed-autoregressive, moving-average model ARIMA (0, 1, 1) implies no autoregressive parameter, one moving-average parameter, and one level of regular differencing A good, though incomplete, introduction to time series modeling is contained in (*9*)

[10] An error structure that is AR(1) is one that exhibits only simple first-order autocorrelation

[11] Adjusted R^2 or $\bar{R}^2 = 1 - (1 - R^2) \{(T - 1)/(T - K)\}$ where T = the number of observations and K = the number of estimated parameters in the regression This adjustment is for degrees of freedom in the estimating equation that have been lost due to the inclusion of additional variables This adjustment offsets the upward bias in the unadjusted R^2 which is most dramatic with a small sample size

What is well documented in the literature, but often overlooked in practice, is that the usual tests of significance, when performed in the presence of autocorrelated errors, are biased

Table 1 presents estimation results which can be considered "nonsense results" (as defined above) as the magnitude of the $\overline{R}^2$ measure exceeds that computed for d (the Durbin-Watson statistic). Table 2 presents results using the same model but employing a simple first-differencing transformation[12] on the dependent and independent series Letting M1* denote the transformed money series, first differencing is accomplished as $M1^*_t = M1_t - M1_{t-1}$ The Durbin-Watson statistic resulting from this estimation is higher than before but it is still very low, a fourth order autoregression[13] was performed on the residuals The results of this autoregression indicate that the error structure is more complex than first-order autoregressive due to significant coefficients on the lagged terms of a lag greater than one [14]

Table 1—Regression results of $ln(CPI_t) = f(ln(M1_t))$

Variable	Estimated coefficients	t-statistic
Intercept	-0 2878	1 42
ln (M1)	9686	[1]21 92
Linear trend	- 0014	[1]3 64

Where R^2 = 0 98
d = 0 06
F = 6027
SEE = 0.1771

[1]Significant at the 0 05 level

Table 2—Regression results of $ln(P_t) - ln(P_{t-1}) = f(ln(M1_t) - ln(M1_{t-1}))$[1]

Variable	Estimated coefficients	t-statistic
Intercept	0 0024	1 61
$ln(M1_t) - ln(M1_{t-1})$	1824	1 53
Linear trend	0001	[2]3 47

Where R^2 = 0 15
d = 0 69
F = 21 7
SEE = 0 0081

[1]Approximates percentage change in price = f (percentage change in M1) Technically, the intercept term should have been omitted, as the intercept values (a vector of ones) have been adjusted by $1 - \rho$, which equals 0 if $\rho = 1$
[2]Significant at the 0 05 level

This example clearly demonstrates how sensitive the summary statistics of a regression can be to a first-differencing transformation This is important both in testing the theoretical specification of the model as well as in forming expectations of its forecasting ability For example, if one obtains an $\overline{R}^2$ of 0 99 from a model estimated by using levels, which is actually explaining only 30 percent of the variation in the dependent variable, the forecasting performance would fall far short of one's expectations Yet, this "model" may be chosen over one estimated using first-differenced data because the $\overline{R}^2$ obtained in the former greatly exceeds that of the latter The choice of the appropriate forecasting model between one estimated from levels and another from changes based on R^2 makes sense only if the R^2 measure truly attests to the model's explanatory power If the model specification is correct, the choice will not matter—they will be identical Presenting these results from raw data and not presenting the Durbin-Watson statistic would be deceptive at best and intellectually dishonest at worst One must also recognize that a "good" Durbin-Watson statistic is insufficient evidence upon which to conclude that the error structure is "contamination free" in terms of autocorrelation, since it tests only for the presence of first-order autocorrelation

Methods of Correcting for Autocorrelation

Assuming that evidence of first-order autocorrelation exists, one then asks what can be done to correct for it, or if correction is appropriate A rather simple (but often effective) approach suggested earlier is to first-difference the data prior to estimation This is equivalent to applying a general filter $(1 - \rho L)$, where L is a lag operator, so that $(1 - \rho L)X_t = X_t - \rho X_{t-1}$ and $\rho = 1$ This technique might alternatively be referred to as "pre-whitening" the input series, "filtering" the input series (or more correctly "pre-filtering" the input series), or "applying a first-difference transform" to the input series Another possible method of eliminating first-order autocorrelation from the error structure is quasi-differencing," or applying the filter $(1 - \rho L)$, where $-1 < \rho < 1$ For example, if ρ is assumed to equal 0 75, applying the filter $(1 - 0\ 75L)$ to Y_t results in the transformed series Y* in which $Y^*_t = Y_t - 0\ 75Y_{t-1}$ The Cochrane-Orcutt iterative technique[15] estimates a value of ρ

[12]First differencing of natural logarithms approximates a percentage rate of change

[13] $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3} + \rho_4 e_{t-4} + u_t$

where $E(e_t u_t) = 0$

[14]For a detailed explanation of one methodology for removing autocorrelation up to and including the fourth-order, see (2, pp 11-14)

[15]Initial and subsequent values of ρ are estimated by a first-order autoregression of the residuals resulting from an OLS estimation using the untransformed data

$$e_t = \rho e_{t-1} + u_t$$

where $E(e_t u_t) = 0$

using the residuals computed from a regression of the untransformed, or "raw," data This technique can choose a value of ρ to satisfy the selection criterion of the computer algorithm but not eliminate the first-order autocorrelation in the error structure This occurs when the estimation converges to a value that is a local, rather than a global, minimization of the sum of squared residuals Another consideration arises if forecasts are generated from a model estimated with this technique as (1) the coefficient of the intercept and its standard error must be corrected,[16] and (2) errors in the forecast will tend to compound over time [17]

Another technique available to the researcher is the Hildreth-Lu procedure, which uses values of ρ prechosen by the researcher This technique is more robust against converging to an inappropriate value of ρ, but it is still vulnerable to the above two considerations if estimates derived from it are used for forecasting

Analysis of Regression Residuals

Referring to the treatment of residuals in econometric models, Granger and Newbold state that

> The traditional approach has been to assume first the residuals to be white noise[18] and to check this assumption by looking at the Durbin-Watson statistic, which effectively measures first-order serial correlation of the residuals If evidence of significant first-order serial correlation is found, the residuals are assumed to be first-order autoregressive there is little reason to suppose that the correct model for residuals is AR(1), in fact if the variables involved are aggregates and involve measurement error, an ARMA[19] model is much more likely to be correct (*13*, p 9)

If the errors are characterized by a mixed, autoregressive, moving-average structure, time series modeling of the residuals can be employed following the methodology of Box and Jenkins (*3*) Although discussion of this technique is beyond the scope of this article, the researcher should be aware of this powerful and innovative approach [20]

Wallis found that models which use quarterly data are often plagued by fourth-order autocorrelation, either when seasonally adjusted or when unadjusted data are used He suggests that monthly or weekly models may also have a seasonal component remaining that can appear either as 12th- or 52nd-order autocorrelation in the error structure, respectively (*27*) We would not suggest that 12th- or 52nd-order prefilters should be constructed and applied to the data, but the residuals can be modeled, again using the techniques developed by Box and Jenkins (*3*) in which one can employ seasonal differencing to the residuals and then estimate the order of the autoregressive and moving-average components of the characterization However, a more detailed discussion of this technique is beyond our scope here

Pierce examines the issue of complex error structures, emphasizing the skepticism that experienced researchers exhibit when confronted with high R^2 measures (*22*) He notes that the R^2 measure is properly constructed as a measure of effects between variables, whereas in many applications, the measure is contaminated by within-variable effects This is seen when lagged values in a relationship are combined with serial correlation in the dependent variable, which "means that part of the variance of y (the dependent variable) is explainable by its own past This R^2 will generally include effects attributable simply to lagged values of y" (*22*, p 3) For this reason, Pierce asserts that the estimated R^2's using time series data exhibit much sensitivity to "prefiltering," which removes this within-variable effect [21] Discussing the great difference between the apparent contribution to R^2 made by a lagged dependent variable expressed as a level (untransformed lagged dependent variable) and the contribution of that variable expressed as changes (that is, first differenced data), Pierce states, "this phenomonon results in an intrinsic ambiguity in conven-

[16] The estimate of the coefficient on the regular intercept must be multiplied by $1/(1 - \rho)$ as must the standard error of that coefficient The other estimated coefficients and the residuals are not affected

[17] This is seen when we consider a forecast generated for one period into the future for some variable Y

$$Y_{t+1} = a_1(1 - \rho) + a_2(X_{t+1} - \rho X_t) + \rho Y_t$$

If Y_{t+1} is understated, Y_{t+2} will also be understated as Y_{t+1} enters the computation on the right-hand side

[18] White noise implies that all the non-random components have been removed from the series and no additional "information" remains

[19] *A*utoregressive *M*oving *A*verage model needing no regular differencing

[20] This procedure attempts to explain the effects of variables which had been excluded from the model One can argue that the parameter estimates thus obtained are more "proper " The forecasting performance is encouraging

[21] The example here does not explicitly include a lagged dependent variable, however, the sensitivity of our example to prefiltering is obvious

The variables included in the final model may be statistically irrelevant, or statistically significant variables may be excluded due to biased t-statistics To protect themselves from being misled by their results, researchers must critically examine the error structure

tional R^2 measures, and it is perhaps this ambiguity which underlies the rather limited faith often accorded these measures by persons experienced with time series data" (22)

Conclusion

Researchers are faced with considerable difficulty in interpreting the results of their modeling efforts Measures computed from summary statistics, such as the coefficient of variation of the dependent variable (standard error of the estimate divided by the mean of the dependent series), are potentially meaningless The explanatory power of the model based on the coefficient of determination may be grossly over- or understated The variables included in the final model may be statistically irrelevant, or statistically significant variables may be excluded due to biased *t*-statistics To protect themselves from being misled by their results, researchers must critically examine the error structure The Durbin-Watson statistic should always be examined[22] as a bare minimum Concurrently, researchers must be aware that this may not be sufficient to insure a meaningful interpretation of their results Additional work in residual analysis is continuing with sophisticated techniques, such as time series modeling, and with analyses in the frequency domain that employ spectral techniques There are no easy answers, researchers must address themselves to some of the complex issues described here if their research is to be useful to policymakers and to other researchers

References

(1) Berndt, E R , and N E Savin "Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances," *Econometrica*, Vol 43, Nov 1979, pp 937-57

(2) Bishop, R. V "An Empirical Investigation into the Causal Relationship Between Money and Prices in the Post World War II United States " Staff report, U S Dept Agr , Econ Stat. Coop Serv , Intnatl Econ Div., June 1980 Paper to be presented at the Midwest Economics Association Meeting, Louisville, Ky , Apr 2-4, 1981

(3) Box, G E P , and G M Jenkins *Time Series Analysis* San Francisco Holden-Day, 1976

(4) ________, and D A Pierce "Distribution of Residual Autocorrelations in Autogressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, Vol 65, Dec 1970, pp 1509-26

(5) Cochrane, D , and G H Orcutt "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, Vol 44, No 245, Mar 1949, pp 32-61

(6) Durbin, J "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, Vol 38, May 1970, pp 410-21

(7) Engle, Robert F "Specification of the Disturbance for Efficient Estimation," *Econometrica*, Vol 42, Jan 1974, pp 135-46

(8) Fuller, W A , and J E Martin "The Effects of Autocorrelated Errors on the Statistical Estimation of Distributed Lag Models," *Journal of Farm Economics*, pp 71-82

(9) Hammond, Elizabeth "Forecasting the Money Stock with Time Series Models " Research Paper 7729 Federal Reserve Bank of New York, May 1977

(10) Kmenta, Jan *Elements of Econometrics* New York Macmillan, 1971

(11) Granger, C W J , and M J Morris. "Time Series Modelling and Interpretation," *Journal of the Royal Statistical Society*, A, Vol 139, Pt 2, 1976, pp 246-57

(12) ________, and P Newbold "Spurious Regressions in Econometrics," *Journal of Econometrics*, Vol 2, Mar 1974, pp 110-20.

(13) ________ "The Time Series Approach to Econometric Model Building," *New Methods in Business Cycle Research Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, 1977, pp 7-22

(14) Griliches, Zvi "Distributed Lags A Survey," *Econometrica*, Vol 35, 1967, pp 16-49

(15) Henshaw, R C , Jr. "Testing Single-Equation Least Squares Regression Models for Autocorrelated Disturbances," *Econometrica*, Vol 34, July 1966, pp. 646-60

(16) Hildreth, C , and J. Y Lu "Demand Relations with Autocorrelated Disturbances " Technical Bulletin 276 Michigan State Univ , Agricultural Experiment Station, 1960

(17) McLeod, A I "Distribution of the Residual Cross-Correlation in Univariate ARMA Time Series Models," *Journal of the American Statistical Association*, Vol 74, Dec 1979, pp 849-55

[22] If a lagged dependent variable is included, the Durbin-Watson statistic is biased toward 2 and is, therefore, more difficult to interpret If it is very low (near 0) or very high (near 4), the residuals are most likely autocorrelated If the sample size is large, one can construct a modified *h*-statistic to test for autocorrelation in the presence of a lagged dependent variable See (*6*)

(18) Maddala, G S *Econometrics* New York McGraw Hill, 1977
(19) ________, and A S Rao "Tests for Serial Correlation in Regression Models with Lagged Dependent Variables and Serially Correlated Errors," *Econometrica*, Vol 41, July 1973, pp 761-74
(20) Nelson, C R *Applied Time Series Analysis for Managerial Forecasting* San Francisco Holden-Day, 1973
(21) Nerlove, N, and K Wallis "Use of Durbin-Watson Statistic in Inappropriate Situations," *Econometrica*, Vol 34, Jan 1966, pp 235-38
(22) Pierce, D A "R^2 Measures for Time Series," *Special Studies Papers 93* Federal Reserve Board of Governors, Oct 1978
(23) Rao, P, and Zvi Griliches "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Autocorrelated Errors," *Journal of the American Statistical Association*, Mar 1969, pp 253-72
(24) Saracoglu, R "The Maximum Likelihood Estimation of Parameters in Mixed Autoregressive Moving Average Multivariate Models" Staff Paper 20 Federal Reserve Bank of Minneapolis, Apr 1977
(25) Theil, H, and K Clements "Recent Methodological Advances in Economic Equation Systems" Unpublished paper, 1979
(26) Tillman, J A "The Power of the Durbin Test," *Econometrica*, Vol 43, Nov 1975, pp 959-74
(27) Wallis, K F "Testing for Fourth Order Autocorrelation in Quarterly Regression Models," *Econometrica*, Vol 40, July 1972, pp 617-36

In Earlier Issues

It was out of the effort to understand the conditions of agriculture and farm people that agricultural economics developed, and this is the driving force that has maintained the life of the science

Oris V Wells
Vol 5, No 1, Jan 1953, p 2