Journal of Econometrics 35 (1987) 359–374. North-Holland

OLS OR GLS IN THE PRESENCE OF SPECIFICATION ERROR?
An Expected Loss Approach

Jerry G. THURSBY*

Ohio State University, Columbus, OH 43210, USA
University of Michigan, Ann Arbor, MI 48109, USA

Received January 1986, final version received February 1987

Omitted variables in regression analysis can lead to the erroneous conclusion that autocorrelation or heteroscedasticity is present. The common response is to use the suggested GLS procedure, even if it is suspected that the error is a non-zero disturbance mean. The question addressed here is whether one is better off with the GLS or with the OLS estimator when the omitted portion of the regression cannot be incorporated into the regression. Using a loss function this paper relates the seriousness of OLS and GLS loss to identifiable parameters. With consistent estimators of these parameters the researcher can choose between OLS and GLS.

1. Introduction

It is well known (and often cited) that the omission of an explanatory variable(s) or use of an incorrect functional form in a regression that otherwise satisfies the full ideal conditions, can lead to the erroneous conclusion that autocorrelation or heteroscedasticity is present among the disturbances.[1] The common practice, however, is to use some generalized least squares (GLS) technique. The question remains as to whether and under what conditions such a practice is warranted. Sims (1972) noted that attempts to achieve efficient estimates may have perverse effects on approximation error in distributed lag models. His arguments involve approximations of (possibly) infinite lag distributions. In a general context he argued that the 'distance' between the actual and estimated lag distribution depends on autocovariance properties of the independent variables so that it may be undesirable to modify data in a search for efficiency. As an example, Sims notes that use of quasi-differenced data to account for autocorrelated residuals can make estimates worse if approximation error is present. Grether and Maddala (1973) examined the

* I wish to express my appreciation to Richard Anderson, Hashem Dezhbakhsh, David Grether, Richard Jensen, C.A. Knox Lovell, G.S. Maddala, J. Huston McCulloch, Randall Olsen, and Mehmet Ozturk for helpful comments. I remain solely responsible for errors.

[1] See, for example, Judge et al. (1985, p. 329), Kmenta (1971, p. 296), and Maddala (1977, p. 291).

0304-4076/87/$3.50

consistency of GLS estimators when errors in variables lead to the appearance of autocorrelation. They make the same point as Sims that quasi-differencing data in an effort to gain efficiency can increase the error in estimation of regression coefficients. McCallum (1976) compared the consistency of GLS using an autocorrelation correction technique and ordinary least squares (OLS) in the presence of an omitted variable for several cases based on specific assumptions about the relation between the included and excluded variables in a simple regression context. He also found that quasi-differencing data can lead to greater estimation error.

This paper builds upon the previous work in three ways:

(i) We use a loss function in analysis of the choice between OLS and GLS. By taking expectations of a loss function with respect to the unidentified parameters of the model we can relate the size of the loss under OLS and GLS to the identifiable parameters of the model. This provides a basis on which researchers can choose between OLS and GLS in particular settings.
(ii) We consider multiple regressions. It is shown in Thursby (1985a) that the loss associated with OLS in the presence of omitted variables rises sharply as the collinearity between *included* regressors increases. The same will be seen to hold for the GLS loss. Plosser (1981) extended Grether and Maddala's (1973) analysis to more than one independent variable and found more ambiguity in multiple regressions than in simple regressions.
(iii) We also consider a case wherein the omitted variable or incorrect functional form suggests that heteroscedasticity is present as well as when the error suggests the presence of autocorrelation.

An overall question addressed by this paper is whether the researcher is better off or worse off with the OLS estimator than with the GLS estimator. Ideally, of course, the researcher would incorporate the omitted portion of the regression directly into the estimation process. This is not always possible; for example, the problem may be that some regressor is unobservable and no reasonable proxy exists. What we wish to ascertain here is whether OLS or GLS should be used conditional on the omitted portion of the regression *not* being directly incorporated into the regression. The OLS/GLS choice is then a choice between the lessor of two evils (inconsistent estimators).

Examination of the OLS/GLS trade-off requires that structure be placed on the regressors of the model. We consider two models examined by the author in Thursby (1985a). The first is a model that is fairly representative of time series regressions and the other is fairly representative of regressions on cross-sections of data. The issue addressed in the earlier paper is the magnitude of OLS loss conditional on parameters of the included regressor covariance matrix. Here we compare that loss with the GLS loss.

In the next section a general outline of the procedure is presented, following which the time series and cross-section models are considered.

2. A general regression model

Consider the model

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + Z_t'\alpha + u_t, \qquad t = 1, 2, \ldots, T, \tag{1}$$

where the regressors X_{1t} and X_{2t} are scalars, Z_t is a column vector of regressors, and t refers to the observational unit. For expositional ease define $X_{3t} \equiv Z_t'\alpha$. Without loss of generality, all variables are assumed to have zero means. β_1, β_2 and α are composed of unknown regression coefficients, and the u_t are independent and identically distributed with mean zero and variance σ_u^2. Let X_{it} $(i = 1, 2, 3)$ be stochastic and independent of u_t and let the vector $X_t = (X_{1t}, X_{2t}, X_{3t})'$ be identically distributed across t with covariance matrix Σ. The researcher is assumed to erroneously *omit* X_{3t} from the regression either because it is unobservable or because the researcher is unaware that it belongs in the model (1). The omission of X_{3t} can refer to either omitted variables or incorrect functional form [in which case X_{3t} represents the sum of second- and higher-order terms in the Taylor series expansion of the true function (if it is analytic)].

The covariance matrix Σ is a function of both an identified parameter vector δ of order m (e.g., it includes the correlation between X_{1t} and X_{2t}) and an unidentified parameter vector γ of order n (e.g., it includes the correlation between X_{1t} and X_{3t}). Depending on the specification of Σ the researcher can be led erroneously to the conclusion that heteroscedasticity or autocorrelation is present in the disturbance u_t.

In general the omission of X_{3t} will cause the OLS and GLS estimators of β_1 and β_2 to be inconsistent (as well as biased). Concentrate on estimation of β_1 and consider the loss function[2]

$$L(\delta, \gamma) \equiv \left(\operatorname{plim} \hat{\beta}_1 - \beta_1\right)^2, \tag{2}$$

for some estimator $\hat{\beta}_1$. Loosely speaking, define the loss in estimation of a coefficient to be the squared systematic error.

As noted in section 1 we wish to relate the seriousness of the inconsistency to the identifiable parameters δ. Denote the prior conditional density of γ as

[2] We choose the function L rather than more commonly used loss functions such as the squared error loss function $(\hat{\beta}_1 - \beta_1)^2$ because we wish to concentrate on the systematic error in estimation of β_1 and ignore the variance of $\hat{\beta}_1$.

$p(\gamma|\delta)$ and the expectation of L with respect to the unidentified parameters is

$$\mathrm{E}[L(\delta,\gamma)|\delta] = \int_{\gamma_1} \dots \int_{\gamma_n} L(\delta,\gamma)\, p(\gamma|\delta)\, \mathrm{d}\gamma_1 \dots \mathrm{d}\gamma_n.$$

For a suitable choice of the prior density this can be evaluated conditional on various values of δ, the vector of identified parameters.

We now turn to specifications of Σ which will generally lead the researcher to conclude that u_t, the disturbance in (1), has a non-scalar covariance matrix. The question is whether the expected loss, for various values of δ, from using OLS is greater or less than the expected loss from using whatever GLS estimator is suggested by the omission of X_{3t}. Define L(OLS) and L(GLS) to be the loss associated with OLS and GLS, respectively. For each alternative specification of Σ we shall consider

$$R \equiv \frac{\mathrm{E}[L(\mathrm{OLS})|\delta]}{\mathrm{E}[L(\mathrm{GLS})|\delta]}.$$

3. A time series model

We use the regression model (1) and add the further assumption that the X_{it} $(i = 1, 2, 3)$ follow the multiple time series model

$$\begin{aligned} X_{1t} &= \rho_1 X_{1t-1} + \varepsilon_{1t}, \\ X_{2t} &= \rho_2 X_{2t-1} + \varepsilon_{2t}, \\ X_{3t} &= \rho_3 X_{3t-1} + \varepsilon_{3t}. \end{aligned} \tag{3}$$

The $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$ are independent, identically distributed random vectors with mean vector zero and covariance matrix

$$\mathrm{E}(\varepsilon_t \varepsilon_t') = \{a_{ij}\}, \qquad i, j = 1, 2, 3,$$

where

$$a_{ii} = \sigma_i^2(1 - \rho_i^2),$$

and

$$a_{ij} = \sigma_i \sigma_j \gamma_{ij}\left((1 - \rho_i^2)(1 - \rho_j^2)\right)^{1/2}.$$

σ_i^2 is the variance of X_{it} and γ_{ij} is the simple correlation between ε_{it} and ε_{jt}. To preserve positive definiteness of $\mathrm{E}(\varepsilon_t\varepsilon_t')$ it is necessary that

$$1 + 2\gamma_{12}\gamma_{13}\gamma_{23} - \gamma_{12}^2 - \gamma_{13}^2 - \gamma_{23}^2 > 0. \tag{4}$$

Consider the OLS estimator of β_1. For the loss function (2)

$$L(\mathrm{OLS}) = \frac{\sigma_3^2}{\sigma_1^2}\left[\frac{\rho_{13} - \rho_{23}\rho_{12}}{1 - \rho_{12}^2}\right]^2, \tag{5}$$

where ρ_{ij} is the simple correlation between X_{it} and X_{jt} and

$$\rho_{ij} = \frac{\gamma_{ij}\left((1 - \rho_i^2)(1 - \rho_j^2)\right)^{1/2}}{1 - \rho_i\rho_j}. \tag{6}$$

It is almost universal practice in a time series regression to test for first-order autocorrelation among the disturbances. The most common test is the Durbin–Watson test with test statistic

$$d = \sum_{t=2}^{T}(e_t - e_{t-1})^2 \Big/ \sum_{t=1}^{T} e_t^2,$$

where the e_t are OLS residuals, $e_t = Y_t - \hat{\beta}_1 X_{1t} - \hat{\beta}_2 X_{2t}$. The statistic d is approximately $2(1 - r)$, where

$$r = \sum_{t=2}^{T} e_t e_{t-1} \Big/ \sum_{t=2}^{T} e_t^2.$$

Unless $\operatorname{plim} r = 0$ the probability of rejecting the null hypothesis of uncorrelated disturbances will be greater than the chosen significance level in large samples. In the present circumstance,

$$\rho \equiv \operatorname{plim} r = A/B,$$

where

$$A = \rho_3 - \rho_3(B_1\rho_{13} + B_2\rho_{23}) - B_1\rho_1\rho_{13} - B_2\rho_2\rho_{23} + B_1^2\rho_1 + B_2^2\rho_2 + B_1B_2\rho_{12}(\rho_1 + \rho_2),$$

$$B = \sigma_u^2/\sigma_3^2 + 1 - 2(B_1\rho_{13} + B_2\rho_{23}) + B_1^2 + B_2^2 + 2B_1B_2\rho_{12},$$

$$B_1 = (\rho_{13} - \rho_{12}\rho_{23})/(1 - \rho_{12}^2),$$

$$B_2 = (\rho_{23} - \rho_{12}\rho_{13})/(1 - \rho_{12}^2).$$

In general, the Durbin–Watson test (as well as other tests for autocorrelation) will be biased toward rejection of the null hypothesis of uncorrelated disturbances. The usual response is to use a GLS estimator that is in the form of the least squares estimator in the regression of $Y_t - rY_{t-1}$ on $X_{1t} - rX_{1,t-1}$ and $X_{2t} - rX_{2,t-1}$. This is essentially the estimator suggested by Cochrane and Orcutt (1949). Since we are dealing with asymptotic results, issues of treatment of the first observation and number of iterations are irrelevant.

Like the OLS estimator, the GLS estimator is generally inconsistent and

$$L(\mathrm{GLS}) = \sigma_3^2 C^2/\sigma_1^2 D^2, \tag{7}$$

where

$$C = \rho_{13}\left(1 + \rho^2 - \rho(\rho_1 + \rho_3)\right)\left(1 + \rho^2 - 2\rho\rho_2\right)$$
$$- \rho_{12}\rho_{23}\left(1 + \rho^2 - \rho(\rho_2 + \rho_3)\right)\left(1 + \rho^2 - \rho(\rho_1 + \rho_2)\right),$$
$$D = \left(1 + \rho^2 - 2\rho\rho_1\right)\left(1 + \rho^2 - 2\rho\rho_2\right) - \rho_{12}^2\left(1 + \rho^2 - \rho(\rho_1 + \rho_2)\right)^2.$$

In evaluating the expected loss of the OLS and GLS estimators, the identified parameters are σ_1^2, σ_2^2, γ_{12}, ρ_1 and ρ_2 and the unidentified parameters are σ_3^2, σ_u^2, ρ_3, γ_{13} and γ_{23}. The relation between OLS and GLS expected loss depends on the choice of prior distribution and to be as general as possible we assume no knowledge about the parameters γ_{13} and γ_{23} [with the exception of the constraint (4)]. Such a state is represented by uniform prior distributions (conditional on γ_{12}) wherein all values of the parameters are treated as equally likely. In some problems one may have notions that certain values are more likely than others and a prior distribution to reflect that knowledge is appropriate. We feel that the choice of a uniform prior represents the most common state of knowledge of researchers and the results presented below are quite general. Nonetheless, at the end of this section we examine the robustness of the results with respect to the choice of a uniform prior.

The range of the parameter γ_{13} is $(-1, +1)$ and for γ_{23} the range is

$$\gamma_{12}\gamma_{13} \pm \left(\left(1 - \gamma_{12}^2\right)\left(1 - \gamma_{13}^2\right)\right)^{1/2},$$

which follows from the covariance restriction (4). For ρ_3 we also use a uniform prior distribution but the range is restricted to $(0, 1)$ in conformance with the general observation that economic variables tend to be positively autocorrelated. Note that we assume ρ_3 to be independent of the other parameters and that the γ_{ij} are related only through the covariance restriction.

The choice of prior densities is not as clear for σ_3^2 and σ_u^2, so we condition on them as well as on the identified parameters. This is not of any particular concern since σ_3^2 enters multiplicatively into the OLS loss (5) and in the same fashion into the GLS loss (7) as well as in a ratio with σ_u^2 in the expression for ρ. σ_u^2 enters only through the ratio with σ_3^2. Since our interest lies in the relative magnitude of the OLS and GLS losses, the factor σ_3^2 which premultiplies (5) and (7) will not affect the comparison. Time series regressions typically have signal noise ratios around ten and σ_3^2 is a portion of the 'signal' variance; this will be used as a basis for choices of σ_u^2/σ_3^2.

The expected OLS loss conditional on the identified parameters and σ_3^2 can be analytically derived. The expression is lengthy, complicated and not intuitively revealing, hence it is not presented here and the interested reader is referred to Thursby (1985a). Unfortunately, the expected GLS loss is not analytically tractable and numerical integrations are used. Thus R, the ratio of expected OLS loss to expected GLS loss, is based on an analytically derived value of the expected OLS loss and a numerically derived value for the expected GLS loss.

Before turning to the results, several points need to be made. First, R varies only as ρ_1, ρ_2, γ_{12}, and σ_u^2/σ_3^2 vary. Since the most readily available measure of correlation among included regressors is ρ_{12} (the correlation between X_1 and X_2) rather than γ_{12} (the correlation between the disturbances ε_1 and ε_2), the results are presented in terms of ρ_{12}. Note that certain combinations of ρ_1, ρ_2, and ρ_{12} violate the condition $-1 < \gamma_{12} < 1$, thus such combinations have no corresponding entry in the table of results. Second, we set $\sigma_u^2/\sigma_3^2 = 0.2$, a value we consider to be reasonable. Some exploration with σ_u^2/σ_3^2 equal to 0.0 and 1.0 reveal that the results are not very sensitive to the chosen value. Finally, in the calculations of R we considered all possible combinations of the values 0.0, 0.25, 0.5, 0.75, 0.90, 0.95, and 0.99 for the parameters ρ_1, ρ_2, and ρ_{12}. Positive values for ρ_1 and ρ_2 are used in keeping with the assumption $\rho_3 > 0$, and the results are independent of the sign of ρ_{12}. Only a subset of the complete results are necessary to characterize the relation of R to those parameter values and table 1, part (a), contains the subset of results.

The results are clear. The values take by ρ_2 and ρ_{12} are not important in determining whether $R \gtreqless 1$ and they have hardly any impact on how far R deviates from 1. The key parameter is ρ_1. If ρ_1 is small, then $R > 1$ and GLS is preferred. If ρ_1 is large, then $R < 1$ and OLS is preferred. Further exploration (not presented) reveals that $\rho_1 = 0.6$ is approximately the value for $R = 1$. The actual value varies close to 0.6 depending on the values of ρ_2 and ρ_{12}.

While it is the case that the value of ρ_1 is the primary determinant of the value of R and ρ_2 and ρ_{12} have little impact, such is not the case for the *levels* of the expected OLS and GLS losses. In part (b) of table 1 are presented the values of the expected OLS loss for the parameter combinations used in part (a) and for $\sigma_3^2/\sigma_1^2 = 1.0$. The primary determinant of the level of the expected loss is ρ_{12}.

Finally, the expected value of ρ, conditional on the identified parameter values used in table 1, was calculated and found to vary in the narrow range

Table 1

Time series model.

	(a) $E[L(OLS\|\delta)]/E[L(GLS\|\delta)]$				(b) $E[L(OLS\|\delta)]$			
	$\|\rho_{12}\|$				$\|\rho_{12}\|$			
ρ_1	0.0	0.5	0.9	0.99	0.0	0.5	0.9	0.99
				$\rho_2 = 0.0$				
0.0	1.25	1.25	1.26	1.26	0.22	0.27	0.85	7.52
0.25	1.18	1.18	1.12	—[a]	0.26	0.31	0.83	—[a]
0.5	1.08	1.08	—[a]	—[a]	0.27	0.31	—[a]	—[a]
0.75	0.92	0.91	—[a]	—[a]	0.25	0.22	—[a]	—[a]
0.9	0.73	—[a]	—[a]	—[a]	0.17	—[a]	—[a]	—[a]
0.99	0.44	—[a]	—[a]	—[a]	0.04	—[a]	—[a]	—[a]
				$\rho_2 = 0.5$				
0.0	1.25	1.23	—[a]	—[a]	0.22	0.24	—[a]	—[a]
0.25	1.17	1.17	1.08	—[a]	0.26	0.31	0.66	—[a]
0.5	1.08	1.08	1.08	1.08	0.27	0.33	1.05	9.22
0.75	0.93	0.92	0.87	—[a]	0.25	0.28	0.33	—[a]
0.9	0.74	0.70	—[a]	—[a]	0.17	0.15	—[a]	—[a]
0.99	0.45	—[a]	—[a]	—[a]	0.04	—[a]	—[a]	—[a]
				$\rho_2 = 0.9$				
0.0	1.26	—[a]	—[a]	—[a]	0.22	—[a]	—[a]	—[a]
0.25	1.18	1.09	—[a]	—[a]	0.26	0.22	—[a]	—[a]
0.5	1.08	1.03	—[a]	—[a]	0.27	0.29	—[a]	—[a]
0.75	0.91	0.90	—[a]	—[a]	0.25	0.29	—[a]	—[a]
0.9	0.73	0.73	0.72	0.72	0.17	0.21	0.67	5.87
0.99	0.46	0.52	—[a]	—[a]	0.04	0.02	—[a]	—[a]
				$\rho_2 = 0.99$				
0.0	1.26	—[a]	—[a]	—[a]	0.22	—[a]	—[a]	—[a]
0.25	1.19	—[a]	—[a]	—[a]	0.26	—[a]	—[a]	—[a]
0.5	1.08	—[a]	—[a]	—[a]	0.27	—[a]	—[a]	—[a]
0.75	0.91	—[a]	—[a]	—[a]	0.25	—[a]	—[a]	—[a]
0.9	0.71	0.69	—[a]	—[a]	0.17	0.20	—[a]	—[a]
0.99	0.44	0.43	0.43	0.43	0.04	0.05	0.16	1.45

[a]Parameter combination implies $|\gamma_{12}| \geq 1$.

0.35–0.40. Thus the observed value of ρ appears to offer no information regarding the OLS/GLS choice.

To this point no knowledge has been assumed concerning the unidentified parameters (except $\rho_3 > 0$). This is probably the most common situation, though there are cases when one has strong prior knowledge about the omitted portion of the regression (for example, the unobserved variable 'ability' in an earnings regression). To augment the above results to allow for more informative priors consider the following. Define R^* to be the ratio of OLS loss [given by (5)] to GLS loss [given by (7)], thus it is the ratio of expected losses using degenerate priors for the unidentified parameters. R^* is a function of all

Table 2

Regression results for dependent variable y ($y = 1$ if GLS loss smallest, $y = 0$ if OLS loss smallest).

(a) Coefficients (t-statistics)

Regressor	Coeff.	(t-stat.)	Regressor	Coeff.	(t-stat.)
Constant	0.534	(19.67)			
ρ_1	−0.775	(−16.54)	ρ_1^2	−0.346	(−6.78)
ρ_2	−0.003	(−0.06)	ρ_2^2	0.007	(0.136)
ρ_3	0.619	(5.96)	ρ_3^2	0.320	(3.47)
ρ_{12}	−0.003	(−0.41)	ρ_{12}^2	−0.037	(−2.51)
ρ_{13}	0.00[a]	(0.00)[a]	ρ_{13}^2	0.134	(9.10)
ρ_{23}	0.005	(0.62)	ρ_{23}^2	0.127	(8.86)

(b) Partial effects

Parameter value	Parameter					
	ρ_1	ρ_2	ρ_3	ρ_{12}	ρ_{13}	ρ_{23}
−0.9	—	—	—	0.063	−0.242	−0.224
−0.75	—	—	—	0.052	−0.202	−0.186
−0.25	—	—	—	0.015	−0.067	−0.058
0.0	−0.775	−0.003	0.619	−0.003	0.000	0.005
0.25	−0.948	0.001	0.779	−0.022	0.067	0.069
0.75	−1.295	0.008	1.098	−0.059	0.202	0.196
0.9	−1.399	0.010	1.194	−0.070	0.242	0.235

[a]Positive but negligible.

parameters with the exception of σ_1^2 and σ_2^2. Values of R^* were calculated for all combinations of ρ_1, ρ_2, and ρ_3 in the set (0.0, 0.25, 0.5, 0.75, 0.9) and ρ_{12}, ρ_{13}, and ρ_{23} in the set (−0.9, −0.75, −0.25, 0.0, 0.25, 0.75, 0.9), subject to the covariance restrictions (4) and $|\gamma_{ij}| < 1$. We also set $\sigma_u^2/\sigma_3^2 = 0.2$. Define a qualitative variable $y = 1$ if $R^* > 1$ and $y = 0$ if $R^* < 1$ (cases of $R^* = 1$ are dropped). The variable y was then regressed on the parameters ρ_1, ρ_2, ρ_3, ρ_{12}, ρ_{13}, and ρ_{23} as well as on their squares and a constant term. There are a total of 6554 observations and the R^2 is 0.574. Part (a) of table 2 gives the coefficient and t-statistics of the regression and part (b) gives partial effects on y of a change in each parameter, evaluated at different values for the parameters.

It is observed from table 2 that the contribution of ρ_2 to the value of y (i.e., the sign of $R^* - 1$) is negligible, a finding in keeping with earlier results. Furthermore, it is only the magnitudes and not signs of ρ_{12}, ρ_{13}, and ρ_{23} that matter and their contributions to the value of y are small compared with those of ρ_1 and ρ_3. From the values of the coefficients of ρ_1, ρ_3 and their squares it would appear that the effects of ρ_1 and ρ_3 nearly cancel for $\rho_1 = \rho_3$.

To investigate further this apparent relationship between ρ_1 and ρ_3 we examined 412,396 values of R^* determined by varying the autoregressive parameters ρ_1 and ρ_2 over the set (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99), the autoregressive parameter ρ_3 over the same set excluding 0 (for which the OLS and GLS losses are equal), and the correlation coefficients ρ_{12}, ρ_{13}, and ρ_{23} over the set (−0.99, −0.8, −0.6, −0.4, −0.2, 0.0, 0.2, 0.4, 0.6, 0.8, 0.99), subject to the covariance restrictions. Of the cases with $R^* > 1$ (i.e., GLS loss is smallest), 92.1 percent are cases with $\rho_3 > \rho_1$ and only 0.9 percent have $\rho_1 > \rho_3$. When $R^* < 1$ (i.e., OLS is preferred), $\rho_1 > \rho_3$ in 86.3 percent of the cases and $\rho_3 > \rho_1$ in only 2.5 percent of the cases. We can say that, in general, $\rho_1 > \rho_3$ implies that OLS is preferred to GLS whereas $\rho_3 > \rho_1$ implies that the smaller loss occurs with GLS.

4. A cross-section model

In this section, a cross-section model is presented which tends to imply that heteroscedasticity is present among the disturbances whereas the true problem is an omitted portion of the regression. The choice between OLS and GLS can again be made on the basis of expected loss.

We again use the basic model (1). It is now assumed that the observations are a random sample made at some point in time on T cross-sectional units. The X_{it} $(i = 1, 2, 3)$ are again assumed to be stochastic with mean zero and variance σ_i^2 and we add the assumption

$$\begin{aligned} \mathrm{E}(X_{it}X_{jt})/\sigma_i\sigma_j &= \rho_{ij} \quad \text{for} \quad t = 1, \ldots, T_1, \\ &= \gamma_{ij} \quad \text{for} \quad t = T_1, \ldots, T, \end{aligned}$$

where ρ_{ij} is not necessarily equal to γ_{ij}. The model implies that the observations come from two 'regions' and that the correlations among the regressors are different in the two regions. Two regions are chosen in order to keep the problem tractable. Define $C = T_1/T$; thus proportion C of the observations come from one region and $1 - C$ from the other.

The vector $X_t = (X_{1t}, X_{2t}, X_{3t})'$ is identically distributed across t in the first region with covariance matrix $\Sigma_1 = \{a_{ij}\}$ where $a_{ii} = \sigma_i^2$ and $a_{ij} = \sigma_i\sigma_j\rho_{ij}$, and in the second region with covariance matrix $\Sigma_2 = \{b_{ij}\}$ where $b_{ii} = \sigma_i^2$ and $b_{ij} = \sigma_i\sigma_j\gamma_{ij}$. In order to preserve positive definiteness of Σ_1 and Σ_2 it is necessary that

$$1 + 2\rho_{12}\rho_{13}\rho_{23} - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 > 0, \tag{8}$$

and

$$1 + 2\gamma_{12}\gamma_{13}\gamma_{23} - \gamma_{12}^2 - \gamma_{13}^2 - \gamma_{23}^2 > 0. \tag{9}$$

For the loss function (2), the loss associated with OLS is

$$L(\mathrm{OLS}) = \sigma_3^2 A^2/\sigma_1^2 B^2, \tag{10}$$

where

$$A = (C\rho_{13} + (1-C)\gamma_{13}) - (C\rho_{23} + (1-C)\gamma_{23})(C\rho_{12} + (1-C)\gamma_{12}),$$

and

$$B = 1 - (C\rho_{12} + (1-C)\gamma_{12})^2.$$

For models such as this wherein the observations come from several regions, researchers frequently use the Goldfeld–Quandt (1965) test for heteroscedasticity. The null hypothesis is that the disturbance variance is constant from observation to observation and the alternative is that the disturbance variance is constant within a region but varies across regions. The statistic, G, is the ratio of the estimated disturbance variances from separate regressions on the observations from the two regions (we assume that there are more observations than regressors in each region) and

$$\operatorname{plim} G = (\sigma_u^2/\sigma_3^2 + A_1)/(\sigma_u^2/\sigma_3^2 + A_2),$$

where

$$A_1 = 1 - 2(B_{11}\rho_{13} + B_{12}\rho_{23}) + B_{11}^2 + B_{12}^2 + 2B_{11}B_{12}\rho_{12},$$

$$A_2 = 1 - 2(B_{21}\gamma_{13} + B_{22}\gamma_{23}) + B_{21}^2 + B_{22}^2 + 2B_{21}B_{22}\gamma_{12},$$

$$B_{11} = (\rho_{13} - \rho_{12}\rho_{23})/(1 - \rho_{12}^2),$$

$$B_{12} = (\rho_{23} - \rho_{12}\rho_{13})/(1 - \rho_{12}^2),$$

$$B_{21} = (\gamma_{13} - \gamma_{12}\gamma_{23})/(1 - \gamma_{12}^2),$$

$$B_{22} = (\gamma_{23} - \gamma_{12}\gamma_{13})/(1 - \gamma_{12}^2).$$

In general the Goldfeld–Quandt test will be biased toward rejection of the null hypothesis of homoscedasticity.

The usual response to a significant Goldfeld–Quandt statistic is weighted least squares where the weights are generally based upon the pooled regression using all T observations. Using the estimated coefficients one estimates the disturbance variance for each region; denote these estimates as s_1^2 and s_2^2.

Observations from region 1 are weighted by $1/s_1$ and observations from region 2 are weighted by $1/s_2$. Conversely, observations from region 2 are weighted by s_1/s_2. The GLS estimator is the OLS estimator using the pooled sample of weighted observations and

$$L(\mathrm{GLS}) = \sigma_3^2 A_3^2 / \sigma_1^2 A_4^2, \tag{11}$$

where

$$A_3 = (C\rho_{13} + W(1-C)\gamma_{13})(C + W(1-C))$$
$$- (C\rho_{23} + W(1-C)\gamma_{23})(C\rho_{12} + W(1-C)\gamma_{12}),$$

$$A_4 = (C + W(1-C))^2 - (C\rho_{12} + W(1-C)\gamma_{12})^2,$$

$$W = (C_1/C_2)^{1/2},$$

$$C_1 = \sigma_u^2/\sigma_3^2 + 1 - 2(B_1\rho_{13} + B_2\rho_{23}) + B_1^2 + B_2^2 + 2B_1B_2\rho_{12},$$

$$C_2 = \sigma_u^2/\sigma_3^2 + 1 - 2(B_1\gamma_{13} + B_2\gamma_{23}) + B_1^2 + B_2^2 + 2B_2B_2\gamma_{12},$$

$$B_1 = D_1/D_3,$$

$$B_2 = D_2/D_3,$$

$$D_1 = (C\rho_{13} + (1-C)\gamma_{13}) - (C\rho_{23} + (1-C)\gamma_{23})(C\rho_{12} + (1-C)\gamma_{12}),$$

$$D_2 = (C\rho_{23} + (1-C)\gamma_{23}) - (C\rho_{13} + (1-C)\gamma_{13})(C\rho_{12} + (1-C)\gamma_{12}),$$

$$D_3 = 1 - (C\rho_{12} + (1-C)\gamma_{12})^2.$$

The unidentified parameters of the OLS and GLS loss functions are σ_3^2, σ_u^2, ρ_{13}, ρ_{23}, γ_{13}, and γ_{23}. With the exception of σ_3^2 and σ_u^2, the expected OLS and GLS losses are evaluated using uniform priors for the unidentified parameters. For the reasons cited in section 3, we condition on reasonable values for σ_u^2/σ_3^2. The expected OLS loss is analytically tractable whereas the expected GLS loss is not. Numerical integration can be used to evaluate the GLS loss but this requires a four-tuple integration. Constraints on computer time require a simplifying assumption to reduce the problem to a more manageable triple integral. In what follows we set $\rho_{23} = \gamma_{23}$; thus the correlation of X_{3t} (the omitted portion of the regression) with X_{2t} (the regressor that is not of interest) is the same in both regions.

Table 3

Cross-section model.

		(a) $E[L(OLS\|\delta)]/E[L(GLS\|\delta)]$				(b) $E[L(OLS\|\delta)]$			
	Sign of	ρ_{12}				ρ_{12}			
γ_{12}	$\rho_{12}\cdot\gamma_{12}$	0.99	0.9	0.5	0.0	0.99	0.9	0.5	0.0
		$C=0.25$							
0.99	+	0.87	0.90	0.98	0.98	6.98	1.25	0.21	0.08
0.9	+	0.98	0.86	0.83	0.87	1.08	0.73	0.26	0.13
0.5	+	0.97	0.93	0.86	0.83	0.25	0.24	0.19	0.15
0.0		0.91	0.89	0.87	0.86	0.14	0.14	0.14	0.14
0.5	−	0.82	0.81	0.82	0.83	0.10	0.10	0.12	0.15
0.9	−	0.85	0.86	0.89	0.87	0.04	0.04	0.07	0.13
0.99	−	0.99	1.00	0.99	0.98	0.01	0.01	0.03	0.08
		$C=0.50$							
0.99	+	0.90	0.96	0.95	0.92	5.58	1.02	0.22	0.10
0.9	+	0.96	0.88	0.91	0.92	1.02	0.59	0.20	0.10
0.5	+	0.96	0.91	0.88	0.89	0.22	0.20	0.15	0.11
0.0		0.93	0.92	0.89	0.88	0.10	0.10	0.11	0.11
0.5	−	0.91	0.93	0.91	0.89	0.05	0.06	0.08	0.11
0.9	−	0.98	0.99	0.93	0.92	0.01	0.02	0.06	0.10
0.99	−	1.03	0.98	0.91	0.92	0.00	0.01	0.05	0.10
		$C=0.75$							
0.99	+	0.89	0.98	0.96	0.90	6.98	1.08	0.25	0.14
0.9	+	0.92	0.87	0.93	0.90	1.25	0.73	0.24	0.14
0.5	+	0.99	0.84	0.86	0.87	0.21	0.26	0.19	0.14
0.0		0.99	0.88	0.84	0.86	0.08	0.13	0.15	0.14
0.5	−	1.00	0.90	0.83	0.87	0.03	0.07	0.12	0.14
0.9	−	1.02	0.87	0.82	0.90	0.01	0.04	0.10	0.14
0.99	−	1.02	0.85	0.81	0.90	0.01	0.04	0.10	0.14

Uniform priors are used for ρ_{13}, γ_{13}, and ρ_{23}. For ρ_{23} the range is $(-1, +1)$ and the ranges of ρ_{13} and γ_{13} are, respectively,

$$\rho_{12}\rho_{23} \pm \left((1-\rho_{12}^2)(1-\rho_{23}^2)\right)^{1/2},$$

and

$$\gamma_{12}\rho_{23} \pm \left((1-\gamma_{12}^2)(1-\rho_{23}^2)\right)^{1/2}.$$

The ranges for ρ_{13} and γ_{13} reflect the covariance restrictions (8) and (9).

For σ_u^2/σ_3^2 we tried 0.0, 0.2, and 1.0. The results are not markedly different and in table 3, part (a), are presented results for $\sigma_u^2/\sigma_3^2 = 0.2$. The table contains the values of R, the ratio of expected OLS to expected GLS loss, for combinations of γ_{12}, ρ_{12}, and C. OLS is preferred except when γ_{12} and ρ_{12} are of opposite signs and both are close to 1 in absolute value. When GLS is preferred, the gain over OLS is slight; when OLS is preferred, the gain can be substantial. For completeness, part (b) gives the corresponding expected OLS losses for $\sigma_3^2/\sigma_1^2 = 1.0$. Finally, the expected value of the Goldfeld–Quandt statistic G, conditional on the identified parameter values of table 3, was calculated and found to vary in the narrow range 1.31–1.54. Thus the observed value of G appears to offer no information regarding the OLS/GLS choice.

As we did with the time series model we drop the uniform priors and examine the ratio of OLS loss [eq. (10)] to GLS loss [eq. (11)]. For all combinations of ρ_{12}, γ_{12}, ρ_{13}, γ_{13}, and ρ_{23} in the set $(-0.99, -0.9, -0.5, 0.0, 0.5, 0.9, 0.99)$, subject to the covariance restrictions, we calculated the ratio of losses. A total of 3111 cases were examined and in only 54 (1.74 percent) of the cases did we observe the OLS loss to be greater than the GLS loss.

An alternative cross-sectional model was also considered. In that model the correlation between regressors is constant across regions but regressor variances in one of the regions are smaller than are the variances in the other region. Expected OLS and expected GLS losses are very close. For the sake of brevity the details of this last model are omitted but are available upon request.

5. Conclusion

It is common knowledge that the errors of omitted variables and incorrect functional form can bias tests for autocorrelation and heteroscedasticity in the direction of rejecting the null hypothesis of non-autocorrelation and homoscedasticity. The common response is to use the suggested GLS procedure, even if it is suspected that the error is a non-zero disturbance mean (conditional on the regressors) rather than autocorrelation or heteroscedasticity. This paper relates the seriousness of OLS and GLS loss to identifiable parameters of the regressor covariance matrix using a loss function. On the basis of consistent estimators of these parameters the researcher can choose between the lesser of two evils (inconsistent estimators of regression coefficients). Of course, the researcher would ideally incorporate the omitted portion of the regression directly into the estimation process. This is not always possible; for example, the error may be the omission of an unobservable regressor or the theory on which the regression is based may be insufficiently developed to suggest possible functional forms.

Two general regression models are presented: a time series and a cross-section model. The expected OLS loss is analytically derived for both models, but the expected GLS loss is analytically intractable and numerical integration is necessary. For the cross-sectional model considered here and the loss function (2), OLS is almost always preferred. For the time series model, the loss function (2), and the restriction $\rho_i > 0$ ($i = 1,2,3$): (a) $\rho_1 > \rho_3$ implies that, in general, there is smaller loss with OLS than with GLS. The opposite is implied by $\rho_3 > \rho_1$; and (b) if the researcher has no prior knowledge concerning ρ_3 (other than $\rho_3 > 0$), then expected OLS loss is smaller than expected GLS loss if $\rho_1 > 0.6$. It should be reiterated that these results apply when the researcher's interest centers on a single regression coefficient and loss involves only inconsistency as in eq. (2). It remains unclear if the results carry over to other loss functions and prior distributions or if interest is on prediction or on estimation of the entire coefficient vector. Nonetheless, the results apply to many commonly encountered problems.

In most applications, researchers have models more elaborate than two regressors, thus the results of sections 3 and 4 do not strictly apply. In the more general setting the term $\beta_2 X_{2t}$ in model (1) is replaced by $X_{2t}'\beta_2$, where X_{2t}' is a row vector of regressors and β_2 is a vector of unknown coefficients. The correlation coefficient ρ_{12} becomes the correlation between X_{1t} and $X_{2t}'\beta_2$, and so forth for the other parameters associated with X_2. Unfortunately, β_2 is generally unidentified because of the omission of X_3, hence ρ_{12}, γ_{12}, and ρ_2 are not identified in this case. Fortunately, however, the cross-sectional model suggests the use of OLS almost exclusively and the choice of estimator in the time series model depends primarily upon ρ_1 which is always identified.

The presumption in this paper is that the researcher has a strong prior that the model under study is misspecified. There is a literature which considers tests to discriminate between misspecification and non-scalar covariance matrices [see, for example, Savin and White (1978), Thursby (1981, 1982)] and tests for specification error [see, for example, Hausman (1978), Plosser, Schwert and White (1982), Thursby and Schmidt (1977), Thursby (1985b), Davidson, Godfrey and MacKinnon (1985)]. In the event one accepts misspecification, the results of this paper can give guidance in the choice between OLS and GLS estimates.

References

Cochrane, D. and G.H. Orcutt, 1949, Application of least squares regression to relationships containing autocorrelated error terms, Journal of the American Statistical Association 44, 32–61.

Davidson, R., L.G. Godfrey and J.G. MacKinnon, 1985, A simplified version of the differencing test, International Economic Review 26, 639–648.

Goldfeld, S.M. and R.E. Quandt, 1965, Some tests for homoscedasticity, Journal of the American Statistical Association 60, 539–547.

Grether, D.M. and G.S. Maddala, 1973, Errors in variables and serially correlated disturbances in distributed lag models, Econometrica 41, 255–262.

Hausman, J.A., 1978, Specification tests in econometrics, Econometrica 46, 1251–1272.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl and T.C. Lee, 1985, The theory and practice of econometrics, 2nd ed. (Wiley, New York).

Kmenta, J., 1971, Elements of econometrics (MacMillan, New York).

Maddala, G.S., 1977, Econometrics (McGraw-Hill, New York).
McCallum, B.T., 1976, On estimation assuming non-existant autocorrelation, Australian Economic Papers, 119–127.
Plosser, C.I., 1981, Serial correlation corrections and distributed lag models in the presence of measurement error, Working paper no. 7830 (University of Rochester, Rochester, NY).
Plosser, C.I., G.W. Schwert and H. White, 1982, Differencing as a test of specification, International Economic Review 23, 535–552.
Savin, N.E. and K.J. White, 1978, Estimation and testing for functional form and autocorrelation: A simultaneous approach, Journal of Econometrics 8, 1–12.
Sims, C.A., 1972, The role of approximate prior restrictions in distributed lag models, Journal of the American Statistical Association 67, 169–175.
Thursby, J.G., 1981, A test strategy for discriminating between autocorrelation and misspecification in regression analysis, Review of Economics and Statistics 63, 117–123.
Thursby, J.G., 1982, Misspecification, heteroscedasticity, and the Chow and Goldfeld–Quandt tests, Review of Economics and Statistics 64, 314–321.
Thursby, J.G., 1985a, OLS inconsistency in the presence of specification error: An expected loss approach, Working paper no. 86-6 (Department of Economics, Ohio State University, Columbus, OH).
Thursby, J.G., 1985b, The relationship among the specification tests of Hausman, Ramsey, and Chow, Journal of the American Statistical Association 80, 926–928.
Thursby, J.G. and P. Schmidt, 1977, Some properties of tests for specification error in a linear regression model, Journal of the American Statistical Association 72, 636–641.