Teaching Regressions with a Lagged Dependent Variable and Autocorrelated Disturbances

Asatoshi Maeshiro

A regression model with a lagged dependent variable and autocorrelated disturbances is a standard subject covered in econometrics textbooks. The estimation problem of these models arises from the correlation between the lagged dependent variable and the current disturbance. This violation of one of the most critical assumptions of the classical linear regression model is supposed to make the ordinary least squares (OLS) estimator *unacceptable* because the estimator will be inconsistent. Textbooks discuss alternative estimators and recommend choices that include such consistent estimators as the maximum likelihood estimator, an instrumental-variables estimator, various feasible generalized least squares (GLS) estimators, and others.

Because the criterion by which the OLS estimator is judged to be inferior to these alternatives is an asymptotic property (after all, all the estimators available in practice are biased), one would expect some ensuing discussion on the finite sample properties of these estimators in order to either justify or qualify recommendations. However, this is not the case in most textbooks. Perhaps some authors might never have doubted the inferiority of the OLS estimator because of the firmly ingrained conventional wisdom derived from asymptotic criteria. Or they might have inferred that the OLS estimator should be inferior, based on the results obtained for a static regression model (no lagged dependent variable) with autocorrelated disturbances. In fact, most authors refer to such results in a preceding chapter. Yet, some might be unaware of the past finite-sample studies on the subject. Even Harvey (1990), one of the few authors who discusses finite-sample properties, cites Maddala and Rao (1973) as "(the) only relevant Monte Carlo work for the lagged dependent variable model" (p. 272), when in fact there are several such studies.

Other authors might have decided to not include such a discussion because of the conflicting results that past studies have produced. Judge et al. (1985), another of the few exceptions, aptly characterizes the results of the finite-sample studies as "depend[ing] on the sample size, the parameters . . . and the exogenous variable . . ." (p. 398). In other words, the conflicting results do not easily render themselves to a coherent summary in a textbook section. Yet Malinvaud (1980)

Asatoshi Maeshiro is a professor of economics at the University of Pittsburgh. *The author wishes to thank William Becker, Hirschel Kasper, and three anonymous referees for their helpful comments and suggestions.*

spares a whole section for the discussion of the finite-sample properties of alternative estimators, but ends up with a pessimistic conclusion that "[w]hen the available series are short, . . . direct least squares fitting is often the most reasonable method of estimating the coefficients" (p. 559). This conclusion is at odds with the *impression* conveyed, intentionally or unintentionally, by most textbooks. Which is accurate?

My purpose in this article is to examine and clarify the seemingly conflicting results of the past finite-sample studies by using a unified theory developed in Maeshiro (1994a). By shedding light on the conflicting results, this article should help both teachers and students in elementary and intermediate econometrics courses understand the intricate factors that determine the relative performances of the OLS estimator and the estimators that take account of the autocorrelation in the disturbances. By showing and explaining how an inconsistent estimator (the OLS estimator in this case) can *consistently* outperform consistent estimators in terms of bias for a finite-sample size, this case should also serve as a showcase example for teaching students how, in practice, asymptotic criteria must be applied with care.[1]

I introduce first a pedagogically useful approach to the econometrics of contemporaneous correlation between a right-hand variable and the current disturbance, examples of which include the current model as well as an errors-in-variables model and a simultaneous-equation model. Next, I provide a heuristic explanation as to the peculiar bias properties of the OLS estimator when the estimator is applied to a simple dynamic model with autocorrelated disturbances. Then, I isolate and analyze the effect of autocorrelation when the disturbances are generated by the stationary AR(1) process, summarize the theory presented in Maeshiro (1994a), and illustrate how the theory can explain the seemingly conflicting results of past Monte Carlo studies.

ECONOMETRICS OF CONTEMPORANEOUS CORRELATION—A PEDAGOGICAL EXCURSION

That the OLS estimator of a regression model is undesirable when explanatory variables and the disturbance term of the model are contemporaneously correlated has long been accepted as a major premise by econometricians. The OLS estimator is regarded as undesirable for such a model because it is biased and inconsistent. Kennedy (1992, 135), in his leading guidebook full of intuition and heuristic explanations, succinctly explains why this happens:

> This is because the OLS procedure, in assigning "credit" to regressors for explaining variation in the dependent variable, assigns, in error, some of the disturbance-generated variation of the dependent variable to the regressor with which that disturbance is contemporaneously correlated. Consider as an example the case in which the correlation between the regressor and the disturbance is positive. When the disturbance is higher the dependent variable is higher, and owing to the correlation between the disturbance and the regressor, the regressor is likely to be higher, implying that too much credit for making the dependent variable higher is likely to be assigned to the regressor.

I shall go one step further and quantify this "credit" by using the following simple linear regression model:

$$Y_t = \alpha + \beta X_t + U_t \qquad (t = 1, \ldots, T), \tag{1}$$

where X_t and U_t are bivariate normal with mean $E(X_t) = \eta$ and $E(U_t) = O$, variance $V(X_t) = \sigma_X^2$ and $V(U_t) = \sigma_U^2$, and covariance $COV(X_t, U_t) = \sigma_{XU} \neq 0$. The T vectors $\{X_t\ U_t\}$, $t = 1, \ldots, T$, are independently and identically distributed. These assumptions are made to simplify mathematical analysis so that a student in an intermediate econometrics course can follow. The consequences of weakening these assumptions will be discussed at the end of this section.

For the present analysis, the cause of the contemporaneous correlation does not need to be specified. This makes the analysis more general. The focal point of the econometrics of contemporaneous correlation, by definition, is the relationship between X_t and U_t and its effect on the properties of estimators of the coefficients α and β. The following decomposition of U_t (which is always possible) explicitly specifies this relationship and provides a cornerstone for the present attempt to assess the above effect:

$$\begin{aligned} U_t &= (\sigma_{XU}/\sigma_X^2)(X_{t-\eta}) + S_t \\ &= \rho(\sigma_U/\sigma_X)(X_{t-\eta}) + S_t, \end{aligned} \tag{2}$$

where ρ is the correlation between X_t and U_t, and S_t is simply the difference between U_t and the first term on the right-hand side. By showing the zero covariance between X_t and S_t, a student should be able to prove the independence (or zero correlation if the normality assumption is dropped) of the two terms. Note that S_t is also independent of other Xs *by assumption,* and hence of any function of the Xs for that matter. Also note that S_t has mean zero and the following variance:

$$V(S_t) = \sigma_S^2 = (1 - \rho^2)\sigma_U^2. \tag{3}$$

Using equation (2), the model can be reformulated as

$$Y_t + [\alpha - \rho\eta(\sigma_U/\sigma_X)] + [\beta + \rho(\sigma_U/\sigma_X)]X_t + S_t, \tag{4}$$

where X_t and S_t are independent as noted. This reformulated model highlights the estimation problem of model (1) described by Kennedy. The credit is $\rho(\sigma_U/\sigma_X)$. If ρ (the correlation between X_t and U_t) is positive, "too much credit for making the dependent variable higher is likely to be assigned to the regressor" (Kennedy 1992, 133). Because this effect is one of the two pillars of my analysis, I shall call it the "correlation effect." Note that the variance of the adjusted disturbance S_t is smaller than that of the original disturbance U_t, which again makes sense because part of the variation in U_t is now accounted for by X_t.

First examine the properties of the OLS estimator in the context of model (1), confining oneself to the estimation of the slope coefficient β. Because the estimation of model (1) by the OLS estimator is equivalent to the estimation of model (4) by the same estimator, one sees immediately that the expected value of the OLS estimator of β of model (1) is given by

$$E(OLS) = \beta + \rho(\sigma_U / \sigma_X) + E\{\Sigma[X_t - \bar{X}]S_t / \Sigma[X_t - \bar{X}]^2\} \qquad (5)$$
$$= \beta + \rho(\sigma_U / \sigma_X).$$

The expected value of $\Sigma[X_t - \bar{X}]S_t/\Sigma[X_t - \bar{X}]^2$ is zero because of the independence of S_t of any functions of the Xs, as noted above. It follows that the bias (which is also the inconsistency in this case) of the OLS estimator of β in model (1) is $\rho(\sigma_u/\sigma_x)$, the correlation effect itself. A comparison of equations (1), (2), (4), and (5) clearly reveals the point emphasized by Kennedy: the OLS estimator of β in equation (1) is biased and inconsistent because it cannot separate the effect of X_t on Y_t from that of U_t on Y_t.

If one relaxes the assumption of independence of successive observations of X_t and U_t, S_t will still be independent of X_t but not of all Xs. Hence, equation (5) is equal to the probability limit (not the expected value) of the OLS estimator, and the correlation effect will represent the inconsistency but not the exact bias of the OLS estimator. In this case, the assumption of joint normality of X_t and U_t is irrelevant as long as the stationarity assumption is maintained. If X_t is nonstationary and the correlation effect varies over time, the estimation problem becomes more complicated. A student might wish to investigate why this is so.

The preceding analysis is based solely on the bias criterion. Although I could adopt the mean square error criterion for further analysis, I shall instead turn to the main theme of this article, a regression with a lagged dependent variable and autocorrelated disturbances. For this model, there will be an additional effect that counters the correlation effect, making the OLS estimator perform remarkably well in terms of bias.

PECULIAR BIAS PROPERTIES OF THE OLS ESTIMATOR —A HEURISTIC EXPLANATION

I begin with the following simple model:

$$Y_t = \alpha + \lambda Y_{t-1} + U_t \qquad (t = 1, 2, \ldots, T) \qquad (6)$$

where $-1.0 < \lambda < 1.0$, and the Us are identically and normally distributed with mean zero and variance σ_u^2.

Suppose that the Us are independently distributed, so that the assumptions of the ideal disturbances are met—independently and identically distributed (i.i.d.) normal random variables. Then, Y_{t-1} and U_t are independent and hence uncorrelated. Therefore, the OLS estimator is generally deemed as an appropriate choice because it is consistent. However, as many textbooks point out, the OLS estimator is biased because the successive Ys are not independent and are correlated with lagged Us. (Unlike the effect of contemporaneous correlation, it is difficult to provide a heuristic explanation as to why it is biased.) The approximate bias of the OLS estimator of λ is given by $-(1 + 3\lambda)/T$ (Kendall 1954), or by $-2\lambda/T$ if no intercept exists (White 1961). (Note that the approximate bias is at least twice as large when an intercept is included and λ is positive. This difference has a direct bearing on the relative performance of the OLS estimator and the estimators that take account of autocorrelation in the disturbances.) This effect on the OLS estimator, which I will call the "dynamic effect," separates the current model from the general model discussed in the preceding section. (In equation [1], if U_t and X_t are independent, the OLS estimator is unbiased.) Note that *if* λ *is positive*, which is the general case in practice, *the dynamic effect is negative*.

Now, suppose instead that the Us are serially correlated. Then, U_t and Y_{t-1} are correlated in general and the framework of analysis for autocorrelated Xs and Us discussed at the end of the previous section is applicable (since the successive Ys are correlated). Thus,

$$Y_t = \alpha^* + (\lambda + \rho\sigma_U/\sigma_Y)\, Y_{t-1} + S_t, \tag{7}$$

where $\alpha^* = \alpha\{1 - \rho\sigma_U/[\sigma_Y(1 - \lambda)]\}$ and Y_{t-1} and S_t are independent. Note that the successive Ys are correlated with lagged Ss, and that the correlation effect is the inconsistency of the OLS estimator as noted in the previous section. Note also that the correlation effect is a function of λ. Otherwise, the OLS estimator *appears* to face the same general estimation problem as in the preceding section. However, this turns out not to be the case.

As noted above, when λ is positive, the sign of the dynamic effect is negative. Thus, if the correlation effect is positive (i.e., the correlation between U_t and Y_{t-1} is positive), the two effects will have opposite signs; hence one would expect the two effects to partially or totally offset each other, reducing the bias of the OLS estimator over a certain range of the parameter(s) that define(s) the disturbance-generating process. In other words, the OLS estimator should perform well over such a range precisely because of positive autocorrelation in the disturbances. This conjecture was substantiated, employing several different disturbance-generating processes, for the simple dynamic model (1) by using a Monte Carlo method in Maeshiro (1988) and also for the general dynamic model by using an approximate bias formula in Maeshiro (1994a). From these studies, one can safely conclude that when λ is positive, the tradeoff between the two effects exists for any disturbance-generating process that creates positive correlation between U_t and Y_{t-1}. In the next section, I present a piece of analytical evidence to further support this conjecture.

For the rest of this article, I confine myself to the case of AR(1) disturbances defined below, because this case has been widely applied in practice, examined in many finite-sample studies, and included in many textbooks in conjunction with the discussion of the Cochrane-Orcutt estimator. However, it is important to recognize that any disturbance-generating process assumed in practice is an approximation, but that, as noted above, the existence of the tradeoff between the two effects is not limited to AR(1) disturbances.

APPROXIMATE OLS BIAS AND CORRELATION EFFECTS OF AR(1) DISTURBANCES

Let us assume that the disturbance term is generated by the stable normal AR(1) process:

$$U_t = \phi_1 U_{t-1} + \varepsilon_t. \tag{8}$$

With this disturbance term, the coefficients of model (1) cannot be identified, as is well known; however, one can still trace the roles of the parameters analytically and reveal the two offsetting effects. Because of its simplicity, many textbooks use this case to show the inconsistency or the bias of the OLS estimator (Intrilligator 1978, 186; Malinvaud 1980, 544–45; Theil 1971, 413; and Judge et al. 1988, 577). That the maximum likelihood estimator may yield two global maxima, the direct consequence of unidentifiability of the model, does not preclude the examination of the properties of the OLS estimator.

By adjusting the method adopted by Kendall (1954), one can obtain the following first-order approximate bias of the OLS estimator of λ (dropping all the terms of second and higher order of λ, ϕ_1, and $1/T$) :

$$BIAS(OLS) \approx \frac{-(1+3\lambda)}{T} + \frac{(T-3)\phi_1}{T}. \tag{9}$$

The first term is the dynamic effect derived by Kendall. The second term is an approximate correlation effect, which has the same sign as ϕ_1. Although crude, as noted below, this approximation serves well as a teaching device. That the well-known bias approximation derived by Kendall follows as a special case of this approximation adds a pedagogical value to it.

Equation (9) clearly reveals the opposing forces of the two effects when λ and ϕ_1 are positive. Given $\lambda > 0$, the approximate bias (in absolute value) will first decrease as ϕ_1 increases, become zero at $\phi_1 = (1 + 3\lambda)/(T - 3)$, and then increase. (This pattern is similar to those disclosed in Maeshiro [1988]). However, the range of ϕ_1, over which the magnitude of the approximate bias is smaller than the magnitude of the dynamic effect (i.e., the range over which the OLS estimator performs well precisely because of positive autocorrelation in the disturbances), would be small even for a large value of λ, except when the sample size is very small. Hence, the implication of the opposing effects for applied econometrics would be limited if equation (9) were a good approximation. It turns out that if one retains higher-order terms in approximation, both the dynamic effect and the approximate correlation effect are larger in absolute value. These larger effects in turn will widen the range of ϕ_1 over which the bias of the OLS estimator is smaller than the dynamic effect. Because such a range of ϕ_1 is a good indicator of the relative bias property of the OLS estimator vis-à-vis an estimator that takes account of autocorrelation in the disturbances (for example, the GLS estimator, feasible GLS estimators, the minimum chi-square estimator, and the maximum likelihood estimator), I employ it frequently below and refer to it as the "OLS-superior range."

Because the tradeoff between the dynamic effect and the correlation effect is the focal point of contention, I present and further analyze the correlation effect associated with AR(1) disturbances, CE[AR(1)]:[2]

$$CE[AR(1)] = \frac{(1-\lambda^2)\phi_1}{1+\phi_1^{\lambda}}. \tag{10}$$

By differentiating CE[AR(1)] (which is just the inconsistency of the OLS estimator for this model) with respect to ϕ_1 and λ, one finds that the correlation effect *increases at a decreasing rate with* ϕ_l while it *decreases at an increasing rate with* λ, given positive values of λ and ϕ_1. Such effects are not captured in the crude approximate correlation effect in equation (9), although a closer approximation with higher-order terms does reveal such interaction effects.

WHAT DETERMINES THEIR OLS BIAS PROPERTIES IN GENERAL MODELS?

From the preceding discussion, it is clear that the bias of the OLS estimator of the simple dynamic model is determined by two effects, the dynamic effect and the correlation effect. It turns out that the same is true for general models that contain exogenous variables on the right-hand side:

$$Y_t = \lambda Y_{t-1} + \beta_1 x_{tl} + \ldots + \beta_k x_{tk} + U_t, \tag{11}$$

where $x_{t1}, \ldots, x_{tk}$ are exogenous variables and the Us are normally distributed stationary AR(1) disturbances. However, the factors that affect the dynamic effect are more complicated. These factors (listed below) are mostly extracted from recent analytical studies by Grubb and Symons (1987) and Kiviet and Phillips (1993). Detailed explanation of these factors is found in Maeshiro (1994a).

For the remaining discussion, I assume that both ϕ_1 and λ are positive, making the signs of the two effects opposite (the dynamic effect being negative and the correlation effect positive). If the signs of the two coefficients are different, the two effects will have the same sign and, therefore, will reinforce rather than offset each other, making the OLS bias larger. (For example, in the standard Koyck [1954] distributed lag model, the reduced form $[Y_t = \alpha + \lambda Y_{t-1} + \beta x_t + \varepsilon_t - \lambda\varepsilon_{t-1}]$ will possess negative dynamic and negative correlation effects, given $\lambda > 0$. Hence, the OLS bias should be negative and large in absolute value. Aside from having an MA(1) disturbance, this model does not represent the general model treated in this article because of these reinforcing effects.)

Factors Affecting the Dynamic Effect

D1. The smaller the sample size, the larger the dynamic effect.

D2. The larger the value of $\lambda > 0$, the larger the dynamic effect, provided the variance of the disturbance term is adjusted to keep the population R^2 approximately constant. If the variance of the disturbance term is fixed, the population R^2 increases with $\lambda > 0$, which in turn tends to decrease the dynamic effect. (This effect is a corollary of factor D3.)

D3. The larger the variance of the disturbance term, the larger the dynamic effect.

D4. The more smoothly trended the exogenous variable(s) (i.e., the higher the correlation between the exogenous variables(s) and the linear time trend), the higher the dynamic effect.

D5. For normal stationary AR(1) exogenous variables, $x_t = \gamma x_{t-1} + z_t$, the larger the value of $\gamma > 0$, the larger the dynamic effect, provided the signal-to-noise ratio is controlled. (This factor is closely related to factor D4 in that the larger the value of γ, the higher the probability that the correlation between a generated series and the linear time trend is high. These exogenous variables were used extensively in the past Monte Carlo studies.)

D6. The larger the number of the exogenous variables, the larger the dynamic effect, provided the variables are of similar nature (i.e., similarly trended or non-trended). In particular, whether or not the constant term is included makes a big difference, as the difference between Kendall's (1954) approximation and White's (1961) approximation reveals.

Note that of these six factors, D1, D4, and D6 are observable directly in practice, but D2 and D3 are unknown although estimable. Factor D5 is useful only for understanding many Monte Carlo studies that used AR(1) exogenous variables.

Factors Affecting the Correlation Effect

Here I extrapolate the prior findings to the general case.

C1. The correlation effect increases with ϕ_1 at a decreasing rate.

C2. The correlation effect decreases with the value of $\lambda > 0$ at an increasing rate.

Combining the two sets of factors results in a theory that the larger the dynamic effect and/or the smaller the correlation effect, the wider the OLS-superior range of ϕ_1. If this is true, my theory predicts that such a range is wide when one or more of the following are true: The sample size is small; the value of λ is high; the value of ϕ_1 is low; the exogenous variable(s) is(are) smoothly trended; the number of the exogenous variables is large; and the relative size of the variance of the disturbance term is large (or the population R^2 is small).

As mentioned earlier, the relative bias properties of the OLS estimator and the estimators that take account of autocorrelation in the disturbances can be indirectly inferred from the comparison of the OLS bias and the dynamic effect, that is, by finding the OLS-superior range. This is possible because these estimators attempt to directly or indirectly eliminate the correlation effect but not the dynamic effect. Hence, the question of the biases of these estimators comes down to the effect of the implied transformation on the factors D1 through D6.[3] Above all, if the transformation decreases the population R^2 as ϕ_1 increases, these biases should increase with ϕ_1. One of the crucial effects of the implied transformation is on the variation of the transformed variables when the exogenous variable(s) is (are) smoothly trended and the disturbances are AR(1) with positive ϕ_1. As explained in Kmenta (1986, 314) and Harvey (1990, 192–93), the transformation drastically reduces the variation and hence reduces the amount of information. The retention of the first set of observations in this case can greatly improve the efficiency of the GLS and related estimators. This is particularly true if the intercept term is in the model. The problem in assessing the effects of these factors is that the implied transformation affects not only the nature of the transformed exogenous variables, but also the variance of the transformed disturbances (if ϕ_1 is known, the variance changes from $\sigma_\varepsilon^2/[1 - \phi_1^2]$ to σ_ε^2), thus further changing the population R^2. If the net effect of transformation on the population R^2 is not large, we should expect the OLS bias to be smaller than the GLS bias over a wider range of ϕ_1 when (1) the sample size is smaller, and/or (2) $\lambda > 0$ is larger, and/or (3) the exogenous variable(s) is (are) smoothly trended, and/or (4) the number of exogenous variables is larger.

COMPREHENDING PAST MONTE CARLO RESULTS

Numerous Monte Carlo studies have been conducted to examine the properties of the OLS estimator and estimators that take account of autocorrelation in the disturbances. Most have concluded that the OLS estimator is generally inferior, but a few concluded otherwise. Equipped with the theory presented above, a student should be able to explain every single result in these studies.[4] To save space, I will illustrate how the theory works by summarily explaining the whole results. (A comprehensive account is given in Maeshiro [1994b], which is available from the author upon request.)

I confine the examination to the estimation of λ. All Monte Carlo studies have used the following model:

$$Y_t = (\beta_0) + \beta_1 X_t + \lambda Y_{t-1} + U_t, \tag{12}$$

where X_t is an exogenous variable, fixed or stochastic, and U_t is a stable normal AR(1) disturbance. Some studies omitted the intercept coefficient.

The past studies can be divided into two groups. The first group consists of studies by Wallis (1967), Hong and L'Esperance (1973), Sargent (1968), and Dhrymes (1971, appendix), that used, as the regressor X_t, a normally distributed random sample and/or normal stationary AR(1) exogenous variables as noted in factor D5, or a fixed variable with no smooth upward trend. This group found that the OLS estimator is inferior to an estimator or estimators that take account of autocorrelation in the disturbances in a majority of the cases. The OLS estimator was found superior when the sample size is small (factor D1) and/or λ is large relative to ϕ_1, confirming the roles of factors D2 and C2. Dhrymes (1971) omitted the intercept (β_0) and obtained the results that show worse performance of the OLS estimator than in Hong and L'Esperance (1973) for comparable types and sizes of exogenous variables, confirming the role of factor D6. He also experimented with negative values of ϕ_1 and found that the OLS estimator performs much worse than for any positive value of ϕ_1. This finding is similar to that predicted by the theory in the previous section because both the dynamic effect and the correlation effect are negative, mutually reinforcing rather than offsetting each other.

Overall, in those studies it was concluded that the OLS estimator is not desirable because more cases were found to not favor the OLS estimator. When one does not know why one estimator performs better under one condition and worse under other conditions, such a conclusion is inevitable but *misleading*.

The second group consists of studies by Maddala and Rao (1973) and Maeshiro (1980), that employed both smoothly trended (e.g., U.S. real GNP series) and nontrended exogenous variables and obtained results from the trended variables that were quite different from those discussed above: consistently superior performance of the OLS estimator, confirming the role of factor D4. In particular, both studies revealed that the range of ϕ_1 over which the OLS estimator performs better than an estimator that takes account of autocorrelation in the disturbances increases with λ at an increasing rate, confirming the joint role of factors C2 and D2. Maddala and Rao did not include the intercept. Had they included it, the performance of the OLS estimator should have been even better because of factor D6. Unlike other studies, Maeshiro used the GLS (not a feasible GLS) estimator (i.e., used the true value of ϕ_1 in the variance-covariance matrix of the disturbances) and still found consistently superior performance of the OLS estimator. (Factors D1, D2, D4, C1, and C2 account for these results.) Maeshiro also examined the effect of increasing σ_ε^2 and summarized that "for positive values of $[\Phi_1]$, there are many combinations of λ and $[\Phi_1]$ for which the effects of the increased variance are critical enough to reverse the ranking of the two estimators in favor of OLS, but in no case is the opposite result found" (p. 728), confirming the role of factor D3.

Thus far, I have focused this analysis on the bias properties. I add that the relative mean square error properties follow the relative bias properties closely, that is, the range over which the OLS bias is smaller than its competitor(s) is quite similar to the corresponding range for the mean square error. In other words, the variance of the OLS estimator is not larger than that of its competitor, or at least not large enough to reverse the ranking based on the bias criterion (Maeshiro 1980, Tables 1 and 2, 725, 727; and Maddala and Rao 1973, Table 5, 767).

As an illustration of how this theory can be applied to a more complicated model, I briefly examine a different Monte Carlo study, undertaken by Spencer (1979), that used a system of two equations:

$$Y_{it} = \alpha_i + \lambda_{it-1} + \beta_i x_{it} + U_{it}, \tag{13}$$

$$U_{it} = \phi_i U_{it-1} + \varepsilon_{it}, \qquad (i - 1, 2) \tag{14}$$

where $(\varepsilon_{1t}, \varepsilon_{2t})$ is i.i.d. bivariate normal with mean zero, variances σ_{11} and $\sigma_{22,}$ and covariance σ_{12}. Spencer's main purpose was to examine whether the asymptotically efficient two-step system estimator suggested by Hatanaka (1976) and Dhrymes and Taylor (1976) performs better than a least-squares-based inconsistent system estimator and selected single-equation estimators, including the OLS estimator, an instrumental-variables estimator, and the estimator proposed by Hatanaka (1974). The efficient two-step system estimator, referred to as the HDT method in Spencer's paper, takes account of both auto-correlation and cross-correlation in the two disturbances, using the residuals obtained by applying a consistent estimator in the first step. The inconsistent system estimator also takes account of auto- and cross-correlations in the disturbances but uses the OLS residuals obtained in the first step. Both variables are nontrended; hence, by factor D4, one should expect poor relative performances of the OLS estimator and the least-squares-based system estimator in general except for the cases of large values of λs and/or small values of ϕs. In summary evaluation, Spencer endorses, as is to be expected, the HDT method as a superior estimator in general but adds the following caveat:

> This general conclusion is clouded by two apparent exceptions to this claim of superiority for the HDT method. For models in which the coefficients of the lagged endogenous variables are . . . large . . . or models in which autoregression [i.e., the value of ϕ_1] is weak, the asymptotically efficient HDT procedure appears to be outperformed in modest samples by estimators based on least squares. (1979, 234)

This is what the theory would have predicted. Moreover, had he employed smoothly trended exogenous variables, he would have obtained more "exceptions"; hence they would have ceased to be exceptions. Once one knows the factors that affect the dynamic effect and the correlation effect, the term "exception" loses its meaning in this context.

An examination of the two tables for the exceptional cases (models 5 and 6) (Spencer, 1979, 239–44) further reveals that even the auxiliary parameters, ϕ_1, ϕ_2, σ_{11}, σ_{22}, and σ_{12}, are more efficiently estimated by the least-squares-based estimator. This is not surprising, because better coefficient estimates given by the OLS estimator in the first stage of the system estimation method should yield the OLS residuals that produce better estimates of the auxiliary parameters. In teaching, we emphasize how the OLS residuals will not provide consistent estimates of these auxiliary parameters and, hence, should not be used. This provides a student with another important lesson of how an asymptotic criterion may not work for a finite sample size.

In the context of the present inquiry, the most noteworthy fact, though not noted by Spencer, is that a system method that "properly" takes account of auto- and cross-correlations in the disturbances (meaning "using consistent estimates of the necessary auxiliary parameters") can be inferior even to the OLS estimator (a single-equation estimator that is supposed to be undesirable) when the dynamic effect is large or the correlation effect is small. A student should now be able to explain why this happens.

Some misleading statements or conjectures pertaining to the relative performances of alternative estimators should be mentioned. Puzzled by the unexpected result that the OLS estimator can consistently outperform the ML estimator, Maddala and Rao (1973, 770) conjectured that "in such models, the likelihood functions are rather skewed and instead of using the ML estimate, we should be considering the mean likelihood estimate. . . ." This may be true, but a better explanation is the good performance of the OLS estimator resulting from the offsetting roles of the dynamic and the correlation effects emphasized in this article. Harvey (1990, 273), in his leading time-series econometrics textbook, singles out the omission of the first set of observations as the reason for the poor performance of the ML estimator in Maddala and Rao's study, because the exogenous variable is smoothly trended. As already mentioned, the retention of the first set of observations would have improved the efficiency of the ML estimator but would not eliminate the dynamic effect and, hence, would not qualitatively change the relative performance of the two estimators. This is particularly true if an intercept is included. Maeshiro (1980), on the other hand, blames the detrimental effect of autoregressive transformation on smoothly trended exogenous variables as a culprit for the poorer performance of the GLS estimator. This is an important cause, as explained earlier, but not the only cause. In fact, for $\lambda = 0.9$, the GLS estimator performs worse than the OLS estimator even when the exogenous variable is nontrended (Maeshiro, 1980, Table 2, 727). A proper explanation must include the increasing effect of λ on the dynamic effect, the decreasing effect at an increasing rate of λ on the correlation effect, and their joint effect on the relative performance of the OLS estimator. Above all, the very fact that the OLS bias can be smaller when the disturbances are autocorrelated than when they are independent must be emphasized, that is, the tradeoff effects. It must also be pointed out that the value of $\lambda = 0.9$ is not a magic number. There is no quantum jump in the performance of the OLS estimator relative to its competitors as λ is increased from 0.8 to 0.9.

CONCLUSION

The bias of the OLS estimator of a regression model with a lagged dependent variable and autocorrelated disturbances is determined by two effects, the dynamic effect and the correlation effect. The former is the OLS bias of the coefficient of the lagged dependent variable when the disturbances are i.i.d. normal; the latter is the effect that contaminates the same coefficient when the current disturbance term is correlated with the lagged dependent variable. When both the coefficient of the lagged dependent variable and the autocorrelation coefficient of AR(1) disturbance are positive, the two effects have opposite signs, thus making the OLS estimator perform well in terms of bias.

My principal purpose in this article has been to rectify the unsatisfactory textbook treatment of the finite-sample properties of estimators of regression models with a lagged dependent variable and autocorrelated disturbances. I have provided a unified theory that can account for all past Monte Carlo results. A caveat: I am not necessarily recommending the OLS estimator for such a model when it may appear to be more efficient, because other considerations such as appropriate standard errors must be taken into account when choosing an estimator.

NOTES

1. In fact, Maddala and Rao (1973, 770-71) wondered *why* the maximum likelihood estimator performs consistently worse than the OLS estimator for a certain range of parameter combinations of the model examined in this article. As far as I know, this question has never been satisfactorily answered.
2. I should note, however, that, strictly speaking, the correlation effect equation (10) is the "asymptotic" correlation effect and should be viewed as an asymptotic approximation to the "true" correlation effect for a given sample size. Meanwhile, the second term in equation (9) is a crude finite-sample approximation to the same "true" correlation effect and can be made more accurate by including higher-order terms. Such a distinction is not necessary for equation (4) because of the assumption that X_t, U_t are i.i.d. bivariate normal random variables.
3. The maximum likelihood estimator includes the term $(1 - \phi_1^2)^{1/T}$ that is unrelated to data transformation.
4. The reader may wonder why the studies cited are of the late 1970s and early 1980s. The subject of this article was investigated in these years. Diffused and conflicting results were obtained, but no satisfactory explanation was provided. Instead of trying to find the reason(s) why the OLS esti-

mator can perform better than a consistent estimator that takes account of autocorrelation in the disturbances, econometricians might have judged this performance of the OLS estimator as a finite-sample anomaly and ceased to investigate further (Note 1).

REFERENCES

Dhrymes, P. J. 1971. *Distributed lags: Problems of estimation and formulation.* San Francisco: Holden-Day.

Dhrymes, P. J., and J. B. Taylor. 1976. On an efficient two-step estimator for dynamic simultaneous equations models with autoregressive errors. *International Economic Review* 17 (June): 362–76.

Grubb, D., and J. Symons. 1987. Bias in regressions with a lagged dependent variable. *Econometric Theory* 3 (December): 371–86.

Harvey, A. C. 1990. *The econometric analysis of time series.* 2nd ed. Cambridge: MIT Press.

Hatanaka, M. 1974. An efficient two-step estimator for the dynamic adjustment model with autocorrelated errors. *Journal of Econometrics* 2 (June): 199–220.

———. 1976. Several efficient two-step estimators for the dynamic simultaneous equations model with autoregressive disturbances. *Journal of Econometrics* 4 (May):189–204.

Hong, D., and W. L. L'Esperance. 1973. Effects of autocorrelated errors on various least squares estimators: A Monte Carlo study. *Communication in Statistics* 2:507–23.

Intrilligator, M. D. 1978. *Econometric models, techniques, and applications.* Englewood Cliffs, N.J.: Prentice-Hall.

Judge, G. G., W. E. Griffiths, R. C. Hill, & T. C. Lee. 1985. *The theory and practice of econometrics.* New York: Wiley.

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lutkepohl, & T. C. Lee. 1988. *Introduction to the theory and practice of econometrics.* New York: Wiley.

Kendall, M. G. 1954. Note on bias in estimation of autocorrelation. *Biometrika* 41:403–04.

Kennedy, P. 1992. *A guide to econometrics.* Cambridge: MIT Press.

Kiviet, J. F., and G. D. A. Phillips. 1993. Alternative bias approximations in regressions with a lagged-dependent variable. *Econometric Theory* 9 (March): 62–80.

Kmenta, J. 1986. *Elements of econometrics.* New York: Macmillan.

Koyck, L. M. 1954. *Distributed lags and investment analysis.* Amsterdam, Netherlands: North-Holland.

Maddala, G. S., and A. S. Rao. 1973. Tests for serial correlation in regression models with lagged dependent variables and serially correlated errors. *Eonometrica* 41 (July): 761–74.

Maeshiro, A. 1980. Small sample properties of estimators of distributed lag models. *International Economic Review* 21 (October): 721–33.

———. 1988. OLS as an estimator of a dynamic model with ARMA errors. In *American Statistical Association 1987 Proceedings of the Business and Economic Statistics Section.* Washington: American Statistical Association, 638–42.

———. 1994a. *Bias approximations in regressions with a lagged dependent variable and autocorrelated disturbances.* Pittsburgh: University of Pittsburgh, Department of Economics.

———. 1994b. *Teaching regressions with a lagged dependent variable and autocorrelated disturbances.* Pittsburgh: University of Pittsburgh, Department of Economics.

Malinvaud, E. 1980. *Statistical methods of econometrics.* Amsterdam, Netherlands: North-Holland.

Sargent, T. 1968. Some evidence on the small sample properties of distributed lag estimators in the presence of autocorrelated disturbances. *Review of Economics and Statistics* 50 (February): 87–95.

Spencer, D. 1979. Estimation of a dynamic system of seemingly unrelated regressions with autoregressive disturbances. *Journal of Econometrics* 10 (June): 227–41.

Theil, H. 1971. *Principles of econometrics.* New York: Wiley.

Wallis, K. F. 1967. Lagged dependent variables and serially correlated errors: A reappraisal of three-pass least squares. *Review of Economics and Statistics* 49 (November): 555–67.

White, J. S. 1961. Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrica* 48:85–94.