Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables

Luke Keele
Department of Politics and International Relations
Nuffield College and Oxford University
George Street, Oxford
OX1 2RL UK
Tele: +44 1865 278716
Email: luke.keele@politics.ox.ac.uk

Nathan J. Kelly
Department of Political Science
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-3265
Email: njkelly@email.unc.edu*

March 23, 2005

*A previous version of this article was presented at the 2003 Southern Political Science Meeting. For suggestions and criticisms, we thank Jim Stimson and Neal Beck.

Abstract

A lagged dependent variable in an OLS regression is often used as a means of capturing dynamic effects in political processes and as a method for ridding the model of autocorrelation. But recent work contends that the lagged dependent variable specification is too problematic for use in most situations. More specifically, if residual autocorrelation is present, the lagged dependent variable causes the coefficients for explanatory variables to be biased downward. We use a Monte Carlo analysis to empirically assess how much bias is present when a lagged dependent variable is used under a wide variety of circumstances. In our analysis, we compare the performance of the lagged dependent variable model to several other time series models. We show that while the lagged dependent variable is inappropriate in some circumstances, it remains the best model for the dynamic models most often encountered by applied analysts. From the analysis, we develop several practical suggestions on when and how to appropriately use lagged dependent variables on the right hand side of a model.

1 Introduction

The practice of statistical analysis often consists of fitting a model to data, testing for violations of the model's assumptions, and searching for appropriate solutions when the assumptions are violated. In practice, this process can be quite mechanical - perform test, try solution, and repeat. Such can be the case in the estimation of time series models.

The Ordinary Least Squares (OLS) regression model assumes, for example, that there is no autocorrelation. That is, the residual at one point of observation is not correlated with any other residual. In time series data, of course, this assumption is often violated. One view of autocorrelation is that it is a technical violation of an OLS assumption that leads to incorrect estimates of the standard errors of estimated coefficients. Applied to time series data, then, the mechanical procedure discussed above would consist of estimating an OLS regression model, testing for autocorrelation, and then using Newey-West standard errors.

But there is a second approach that involves thinking of time series data in the context of political dynamics. Instead of worrying about the residuals, we can develop theories and use statistical models that capture the dynamic processes in question. With respect to autocorrelation, we might develop a theory that includes dynamics and correct the specification of the model by making it theoretically appropriate instead of trying to fit a static linear model with the correct standard errors. In other words, analysts should view autocorrelation as a potential sign of improper theoretical specification rather than just a narrow violation of a technical assumption (Beck 1985; Hendry and Mizon 1978; Mizon 1995).

The second solution is far better on the grounds of advancing theories that help us understand the dynamics of politics, and lagged dependent variable models are a statistical tool that aid in this pursuit. In the study of public opinion, for example, we can conceive of theories in which an attitude at time t is a function of that same attitude at $t-1$ as modified by new information rather than viewing an attitude at time t as a linear function of independent variables. Lagged dependent variable models provide a straightforward statistical representation of such a theory. In point of fact, for behavior that we understand to be dynamic decision-making, the appropriate model will also be dynamic. In order to test such dynamic theories, previous attitudes *must* be a component of any plausible statistical model, and any model that omits such a dynamic component is under-specified.

When testing theories that have a dynamic component, then, lagged dependent variable models are more than just theoretically preferable to static models with corrections for autocorrelation. A dynamic process modeled with a static model is invariably misspecified and therefore incorrect. However, the properties of lagged dependent variable models estimated with OLS are not perfect and, worse, these imperfections are not as well understood as they should be. As a result of the uncertainties that surround these models, the lagged dependent variable model is often much maligned. In reality, the problems with lagged dependent variable models are often trivial and confined to situations that are rarely encountered in applied data.

Our aim is to clear up this confusion. In the following sections, we begin by outlining the theory behind lagged dependent variables, and, then, we define precisely the conditions under which problems may arise in the estimation of these models. We perform a Monte Carlo analysis to eliminate the uncertainties that surround lagged dependent variable models and identify conditions under which the estimation of these is most appropriate. We conclude by offering guidelines to applied researchers regarding when a lagged dependent variable model should be estimated and how to ensure that the model is appropriate.

2 The Logic and Properties of Lagged Dependent Variables

In this section, we begin with a conceptual discussion of lagged dependent variable (LDV) models. We discuss the LDV model as a special case of a more general dynamic regression model that is designed to capture a particular type of dynamics. By describing the type of dynamics captured with LDV models, we hope to remind applied analysts of the underlying theory they represent. We then delineate the statistical properties of OLS when used with an LDV and identify where uncertainty exists with regard to the empirical performance of these models.

2.1 The Logic of LDVs

Any consideration of LDV models must start with the autoregressive distributed lag (ADL) model, which is, particularly in economics, the workhorse of time series models. The ADL model is fairly simple and is usually represented in the following form:

$$Y_t = \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t \tag{1}$$

Specifically this model is an ADL(1,1), where the notation refers to the number of lags included in the model and generalizes to an ADL(p,q) where p refers to the number of lags of Y and q refers to the number of lags of X.[1] If $\beta_1 = 0$, then one estimates a lagged dependent variable model:

$$Y_t = \alpha_1 Y_{t-1} + \beta_0 X_t + \varepsilon_t \tag{2}$$

where the only lagged term on the right hand side of the equation is that of Y, the dependent variable.[2]

Why would one estimate such a model? One possible reason is to rid the model of autocorrelation. While in practice the inclusion of a lagged dependent variable will accomplish this, the more appropriate reason for estimating equation (1) is to capture, in a statistical model, a type of dynamics that occurs in politics. For example, theory may predict that current presidential approval is influenced by the current state of the economy. However, theory also dictates that the public remembers the past, and this implies that the state of the economy in previous periods will matter to presidential approval today. To test this basic theory, an analyst might decide to fit the following model:

$$Approval_t = \alpha + \beta_0 Econ_t + \beta_1 Econ_{t-1} + \beta_2 Econ_{t-2} + \beta_3 Econ_{t-3} + \varepsilon_t \tag{3}$$

This model could be estimated, but the lagged X's will undoubtedly be highly collinear, leading to imprecise estimates of the $\beta's$. A better way to proceed would be to assume some functional

[1] For a nice discussion of ADL(1,1) models see Hendry (1995).

[2] Such models are often referred to as partial adjustment models in the econometrics literature.

form for how effects of economic evaluations persist. One could assume that these effects decay geometrically, implying that the state of the economy from the last period is half as important as the current state of the economy, and the economy from two periods ago is half as much again as important. We would then have the following, more general, model, which can be rearranged as follows:

$$
\begin{aligned}
Y_t &= \alpha + \beta_0\lambda^0 X_t + \beta_1\lambda^1 X_{t-1} + \beta_2\lambda^2 X_{t-2} + \beta_3\lambda^3 X_{t-3} + ... + \varepsilon_t \\
&= \alpha + \frac{\beta_0}{1-\lambda L} X_t + \varepsilon_t \\
&= \alpha + \lambda Y_{t-1} + \beta_0 X_t + \varepsilon_t \qquad (4)
\end{aligned}
$$

In this specification, a lagged value of presidential approval captures the effects of the past economy.[3] Although the coefficient β_0 represents the effect of the current economy on current presidential approval (controlling for lagged presidential approval), the effects of past economic performance persist at a rate determined by the autoregressive effect of lagged Y_t. Thus, the effects of X_t will resonate not only in the current quarter but also feed forward into the future at the rate: $\frac{\beta_0}{1-\lambda}$.

One way to describe this specification is to say that presidential approval today is a function of past presidential approval as modified by new information on the performance of the economy. The lagged dependent variable model has a dynamic interpretation as it dictates the timing of the effect of X on Y. This makes it a good choice for situations in which theory predicts that the effects of X variables persist into the future. Furthermore, since autocorrelation can be the result of a failure to properly specify the dynamic structure of time series data, the lagged dependent variable can also eliminate autocorrelation present in a static regression that includes only the current state of

[3]The steps required to go from the second to third part of equation 4 are not entirely trivial in that the mathematics raise an important issue about the error term. The last line of equation 4 is actually the following: $Y_t = (1-\lambda)\alpha + \lambda Y_{t-1} + \beta_0 X_t + u_t - \lambda u_{t-1}$. The non-trivial part of this equation is the error term, $u_t - \lambda u_{t-1}$, which is a MA(1) error term. Most discussions of this model simply note that it is an MA(1) error term and move on. Beck (1992), however, has a nice treatment of this issue and notes that this MA(1) error term can be represented as an AR process (or is empirically impossible to distinguish from an AR process). The nature of the error term as an AR process is important for determining the properties of OLS when used with a lagged dependent variable and is taken up in the next section.

the economy as an explanatory factor. In other words, specification induced autocorrelation can be eliminated when dynamics are appropriately captured with an LDV, making the LDV solution to autocorrelation a theoretical fix for a technical problem in at least some circumstances. We, next, explore the complications that arise when OLS is used to estimate a model with a lagged dependent variable.

2.2 The Properties of OLS With LDVs

Generally, models with LDV's are estimated using OLS, making this an easy specification to implement. The OLS estimator, however, produces biased but consistent estimates when used with a lagged dependent variable if there is no residual autocorrelation in the model (Davidson and MacKinnon 1993). The proof of this appears infrequently, so we reproduce it to help clarify the issues that surround the estimation of LDV models. Consider a simple example where:

$$y_t = \alpha y_{t-1} + \varepsilon_t \tag{5}$$

We assume that $\mid \alpha \mid < 1$ and $\varepsilon_t \sim IID(0, \sigma^2)$. Under these assumptions we can analytically derive whether the OLS estimate of α is unbiased. The OLS estimate of α will be:

$$\hat{\alpha} = \frac{\sum_{t=2}^{n} y_t y_{t-1}}{\sum_{t=2}^{n} y_{t-1}^2} \tag{6}$$

If we substitute (5) into (6) we find that:

$$\hat{\alpha} = \frac{\alpha \sum_{t=2}^{n} y_{t-1}^2 + \sum_{t=2}^{n} \varepsilon_t y_{t-1}}{\sum_{t=2}^{n} y_{t-1}^2} \tag{7}$$

And if we take expectations, the estimate of α is the true α plus a second term:

$$\hat{\alpha} = \alpha + \frac{\sum_{t=2}^{n} \varepsilon_t y_{t-1}}{\sum_{t=2}^{n} y_{t-1}^2} \tag{8}$$

Finding the expectation for the second term on the right hand side of (8) is not easy and, at this point, is unnecessary other than to say it is not zero. This implies that models with lagged dependent variables estimated with OLS will be biased, but all is not lost. If we multiply the right hand side by n^{-1}/n^{-1} and take probability limits, we find the following:

$$\text{plim}_{n\to\infty}\,\hat{\alpha} = \alpha + \frac{\text{plim}_{n\to\infty}(n^{-1}\sum_{t=2}^{n}\varepsilon_t y_{t-1})}{\text{plim}_{n\to\infty}(n^{-1}\sum_{t=2}^{n} y_{t-1}^2)} = \alpha \tag{9}$$

So long as the stationarity condition holds (that is, if $|\alpha| < 1$) the numerator is the mean of n quantities that have an expectation of zero. The probability limit of the denominator is finite, too, so long as the stationarity condition holds. If so, the ratio of the two probability limits is zero and the estimate of α converges to the true α as the sample size increases. As is often pointed out, the finite sample properties of the OLS estimator of $\hat{\alpha}$ are analytically difficult to derive (Davidson and MacKinnon 1993), and investigators must often rely on asymptotic theory.[4]

The key to the above proof is the assumption that the error term is IID. Only then is OLS consistent when used with an LDV. Achen (2000), however, argues that the assumption that $\varepsilon_t \sim IID(0, \sigma^2)$ is not an innocuous one given the consequences of violating this assumption. Importantly, we can violate this assumption under two different contexts. Achen, first, considers violating this assumption when the model is what we call "static" as opposed to dynamic (that is, there are no dynamics in the form of lagged values).[5] To understand what we mean by a static model, we write the following general data generating process (DGP):

$$Y_t = \alpha_1 Y_{t-1} + \beta_1 X_t + u_t \tag{10}$$

where:

$$X_t = \rho_1 X_{t-1} + e_{1t} \tag{11}$$

and

$$u_t = \rho_2 u_{t-1} + e_{2t} \tag{12}$$

[4]That is not to say they are impossible to derive. Hurwicz (1950); Phillips (1977); White (1961) have all derived the small sample properties of α analytically.

[5]What we refer to as a static model is not strictly a static model. A truly static model does not have an autoregressive error term. A model with an autoregressive error but no lags in the data generating process is more correctly termed a common factor or COMFAC model. However, since this model does not include a lag of Y (or any other lags) it is in some sense "static."

For the DGP above, α_1 is the dynamic parameter, and ρ_1 and ρ_2 are the autoregressive parameters for X and the error term of Y respectively. In the discussion, here, we assume that all three processes are stationary, that is α_1, ρ_1, and ρ_2 are all less than one in absolute value. We are in the static context when $\alpha = 0$ and the autoregressive properties of the model are solely due to autocorrelation in the error term. Achen considers the consequences of incorrectly including a lag of Y when the model is static. He demonstrates that the estimates of both α_1 and β_1 will be biased by the following amounts if an LDV is incorrectly included in the estimating equation:

$$\text{plim}\ \hat{\beta} = \left[1 - \rho_1\rho_2\frac{1-R^2}{1-\rho_1^2R^2}\right]\beta_1 \tag{13}$$

$$\text{plim}\ \hat{\alpha} = \alpha + \rho_2\frac{1-R^2}{1-\rho_1^2R^2} \tag{14}$$

Here, R^2 is the (asymptotic) squared correlation when OLS is applied to the correctly specified model (that is one without a lag of Y). Clearly, the estimate of β will be downwardly biased as both ρ_2 and ρ_1 increase. In this static context, imposing a dynamic specification in the form of a lagged dependent variable is a case of fitting the wrong model to the DGP; one where the analyst should not include dynamics where they do not belong. The consequences of an extraneous regressor, in the form of a lag of Y, are worse than normal where we would only expect an increase in the variance of the OLS estimator. Achen recommends the use of OLS without any lags, which will produce unbiased estimates with Newey-West standard errors to provide correct inferences. Achen's critique of the LDV model, however, is more far reaching.

Achen, next, considers the case where the lag of Y is correctly part of the DGP, a situation in which the model actually is dynamic. He derives the asymptotic bias in α_1 and β_1 for the dynamic context:

$$\text{plim}\ \hat{\alpha} = \alpha + \rho_2\frac{\rho_2\sigma^2}{(1-\rho_2\alpha)s^2} \tag{15}$$

$$\text{plim}\ \hat{\beta} = \left[1 - \frac{\rho_1 g}{1-\rho_1\alpha}\right]\beta_1 \tag{16}$$

where $s^2 = \sigma^2_{y_{t-1},x_t}$ and $g = \text{plim}(\hat{\alpha} - \alpha)$. If ρ_2 is zero, then OLS is consistent as demonstrated earlier. However, if $\rho_2 \neq 0$, the estimate of β_1 will be biased downward as ρ_1 and ρ_2 increase (Achen 2000; Griliches 1961; Hibbs 1974; Maddala and Rao 1973; Malinvaud 1970; Phillips and Wickens 1978) just as in the static context. It would appear, then, that if the errors are autocorrelated at all, even in the dynamic context, fitting an LDV model with OLS is problematic and unadvisable.

Before the last rites are read to the LDV model, however, four points must be made. First, for Equation (10) to be stationary, the following condition must be satisfied: $|\alpha + \rho_2| < 1$.[6] If the model is non-stationary, an LDV model is both statistically and theoretically the incorrect model. The LDV implies a process that does not have a permanent memory, once the model is non-stationary the process does have a permanent memory and a different set of statistical techniques apply. But if the DGP is dynamic, we expect the value of alpha to be large (generally above 0.60), which implies that the value for ρ_2, must be small. This limit on the value of ρ_2 implied by the stationarity condition suggests that there can only be minimal amounts of autocorrelation in the residuals if the DGP is dynamic, which should limit the amount of bias.

Second, while OLS without an LDV and a correction for the standard errors may be the best model when α is 0, as soon as α is non-zero, OLS without an LDV will be biased due to an omitted variable. In short, if $\alpha \neq 0$, omitting the lag of Y is a specification error, and the bias due to this specification error will worsen as the value of α increases. Therefore, if we advocate the use of OLS without an LDV and the model is dynamic ($\alpha \neq 0$), we run the risk of encountering bias in another form. Moreover, this bias may be worse than what we might encounter with an LDV model even if the residuals are autocorrelated.

Third, the likelihood of being in the static context is small. Such a model implies that history has no effect on current values of the dependent variable - that the causal process has no memory. Such a model would imply, for example, that past economic performance does not matter to current presidential approval, only current economic performance matters. In short, if we suspect that a process at time t is a function of history as modified by new information, we cannot be in the static context. Moreover, static models of this type have been strongly criticized in the econometrics

[6] This is only roughly true see the appendix for the exact stationarity conditions.

literature (Mizon 1995, 1977) where they are generally thought to be an unrealistic statistical representation of empirical time series processes.

And finally, the analytical results in the dynamic context must rely on values of s^2 that can only be set arbitrarily. While we can pick a range of values for s^2, in reality we need experimental evidence where the values of s^2 vary as they might with real data.

In short, in the real-world of hard choices, it matters a great deal whether we are facing an expected bias of 0.001 or 0.01 or even a larger value that fundamentally alters the magnitude of the coefficients and the inferences we make with them. The analytic results imply bias, but do not give us practical guidelines about the magnitude of this bias given the above considerations. Moreover, with an experimental analysis, we can compare several estimators that are plausible, but for which analytical results are much more difficult to derive. What follows, then, is a Monte Carlo study of various estimators in two different contexts. In the first context, we assume that the bulk of the autocorrelation is in the error term while incorrectly including a lag of Y. In the second context, we include the lag of Y in the DGP such that the model is dynamic, but we vary ρ_2 such that the error term is autocorrelated rendering OLS inconsistent.

3 Monte Carlo Analysis

The analytic results suggest that we study how the LDV model behaves in both the static and dynamic contexts compared to other plausible models and estimators. To that end, we start with the same general DGP from above:

$$X_t = \rho_1 X_{t-1} + e_{1t} \tag{17}$$

$$Y_t = \alpha_1 Y_{t-1} + \beta_1 X_t + u_t \tag{18}$$

$$u_t = \rho_2 u_{t-1} + e_{2t} \tag{19}$$

where e_{1t}, $e_{2t} \sim IID(0, \sigma^2)$.

In the initial experiments, we set ρ_1 to 0.95 to make the test against the LDV as stringent as possible. For as can be seen in the previous section, higher amounts of auto-regression in X also has an adverse impact on the estimates of the LDV model. In the first experiment, we fix ρ_2 to 0.75 and we study the situation in which an LDV model is fitted to a DGP ranging from static ($\alpha_1 = 0.0$) to what we will call weakly dynamic ($\alpha_1 = 0.10$ or 0.20). In the second experiment, we create a dynamic DGP. Here, we fix α to 0.75, while ρ_2 is set to 0.00, 0.10, or 0.20. In both experiments β_1 is set to 0.50. In the first experiment, we (incorrectly) include the lag of Y in the estimating equation, while in the second experiment, we compare the results of estimating the LDV model in the presence of autoregressive error to other models and estimators.

In both experiments, we fit several models. The first is an OLS LDV model. Next, we fit an ARMA(1, 0) model estimated via MLE. This model implicitly includes an LDV, but imposes a different theoretical structure on the data. Third, we include a model fitted with GLS; specifically, we estimate a Cochrane-Orcutt regression.[7] Finally, we estimate an OLS model without an LDV. In all these experiments, we fix N to be 250. Each Monte Carlo experiment was repeated 1000 times.

For each model, we record both the root mean square error[8] (RMSE) and the bias in the estimates of β_1, which we report as a percentage. The performance of OLS with and without an LDV should move in exactly opposite directions. As the model becomes more dynamic (as the value of α_1 increases) the performance of OLS without an LDV should decline. The performance of OLS with an LDV, however, should improve as the value of α_1 increases, but the gains in performance may be minimal if the errors are autocorrelated.

[7]We also considered reporting the results of estimating the LDV model with two-stage least squares using two lags of X as instruments for the LDV. We felt that this test is too friendly to the two-stage least squares estimator, since we can easily construct the model such that past lags of X, the instruments, are not uncorrelated with past lags of the error term. However, as Bartels (1991) notes, this is not a realistic assumption.

[8]More specifically, the RMSE is: $\sqrt{\sum_{l=1}^{1000}(\hat{\theta} - \theta)^2}$. A measure of both bias and variance. The RMSE calculation, here, only includes β_1 the parameter associated with the X variable.

4 Results

4.1 The Static Model

We, first, report the results for the static to weakly dynamic DGP. We begin with the RMSE for all the models. It is immediately clear that the LDV model performs much worse than all the other methods of estimation under this set of experimental conditions. The RMSE for the LDV model is around 0.30 while for the other models the RMSE was typically below 0.10. The RMSE for the LDV model is the same regardless of whether the DGP is static or weakly dynamic. The ARMA model has the lowest RMSE; however, the difference between the ARMA model and the GLS model is trivial (on average 0.05 for the ARMA model and 0.06 for the GLS model).

Table 1: RMSE For $\hat{\beta}$, The Coefficient Of X

	α		
Estimator	0.00	0.10	0.20
LDV	0.304	0.307	0.307
ARMA	0.055	0.059	0.068
GLS	0.055	0.060	0.071
OLS	0.075	0.099	0.152

Results are based on 1000 Monte Carlo replications. ρ_2 : 0.75

The effect of the omitted variable bias in the OLS model without an LDV is clearly discernible in the RMSE. While the RMSE is quite low when the DGP is static, the RMSE doubles from 0.075 when the model is static to 0.152 once α is 0.20. We, next, report the amount of bias in the estimate of β_1. Table 2 reports the bias in the estimates of β_1 as percentage to help the reader understand more clearly the effect of using an LDV in the static or weakly dynamic context.

Under the static DGP, the bias is minimal for all the models except the LDV model where the bias is considerable. If an LDV is wrongly included in a static model, β_1 will be underestimated by just over 60%, while the bias for the other models is typically well under 1%. The bias increases across all the other estimators as the model becomes increasingly dynamic. The effect of the omitted variable bias is particularly noticeable for the OLS model. The bias is under 1% for the

Table 2: Percentage Of Bias In $\hat{\beta}$, The Coefficient Of X

Estimator	α		
	0.00	0.10	0.20
LDV	−60.21	−60.74	−60.78
ARMA	−0.05	3.02	5.08
GLS	−0.07	3.51	6.32
OLS	−0.62	10.68	23.74

Results are based on 1000 Monte Carlo replications. Cell entries represent the average biases in estimated coefficient. ρ_2 : 0.75

static model, but is a sizeable 24% once the model is weakly dynamic. The bias, here, is in the opposite direction from that of the LDV model as $\hat{\beta}_1$ is too large. The consequences of fitting a dynamic model in the form of an LDV to static data are abundantly clear. However, we shouldn't be surprised that fitting the wrong model for the data generating process produces biased results. We, next, perform the same experiment for a dynamic DGP.

4.2 Dynamic Results

With a dynamic DGP, history matters, as past values affect current values of Y. More specifically, α is now 0.75, while ρ_2 is 0.0, 0.1, or 0.20. The critical test, now, is how does the performance of the LDV model change when ρ_2 is above 0.0. We start with the RMSE results for when there is no residual autocorrelation in the error term. Under this condition, we expect OLS with an LDV to be consistent. While we cannot assess whether the fit of the LDV model is improving as the sample size increases, since the sample size is fixed, the LDV model performs well as it has the lowest RMSE of all the models estimated. The RMSE for the LDV model is much lower than that of OLS without and LDV and the GLS model (0.04 compared to 1.2 and 0.52, respectively). The question, though, is will the LDV provide a superior estimates once the error is no longer IID?

Remarkably, the RMSE for the LDV model only increases slightly (from 0.040 to 0.042) once we introduce autocorrelation into the error term. The RMSE increases to a mere 0.050, the lowest of all the models considered, when the autocorrelation in the error term is at its maximum. While the analytic results tell us that OLS is inconsistent, the amount of bias in the estimate of β appears to be trivial.

The GLS model and OLS without an LDV are clearly dominated by the other models. The GLS RMSE is, on average 10 times higher than that of the LDV model. The RMSE for OLS without an LDV is nearly 30 times larger than the LDV model. Clearly, the omission of dynamics from the model induces serious specification bias. The effect of such dynamic misspecification is also evident in the bias for β_1, the coefficient for the X variable.

Table 3: RMSE For $\hat{\beta}$, The Coefficient Of X

	ρ_2		
Estimator	0.00	0.10	0.20
LDV	0.040	0.042	0.050
ARMA	0.076	0.077	0.078
GLS	0.518	0.480	0.439
OLS	1.160	1.162	1.163

Results are based on 1000 Monte Carlo replications. α_1 : 0.75

Table 4 contains the bias, as a percentage, for all the models in the dynamic context. The bias for the LDV model is a minimal two percent when there is no residual autocorrelation. The bias in the LDV model increases slightly as the error becomes more autocorrelated, but the bias is generally smaller than the ML estimates. What is most noticeable, however, is the extremely poor estimates from the GLS and OLS models. For GLS, the estimates are nearly 100% too large, while the OLS estimates are over 200% too large. Clearly, both OLS and GLS are poor choices when the process is dynamic.

The initial evidence suggests that the LDV model is robust to violations of the error term assumptions. In the next section, we further explore the performance of the LDV model under a wider variety of dynamic contexts with closer attention paid to the quality of the inferences we make with the LDV model.

Table 4: Percentage Of Bias In $\hat{\beta}$, The coefficient Of X

Estimator	ρ_2		
	0.00	0.10	0.20
LDV	2.21	−1.29	−5.19
ARMA	−3.60	−4.05	−4.36
GLS	96.42	88.25	79.38
OLS	230.50	230.76	230.91

Results are based on 1000 Monte Carlo replications. Cell entries represent the average biases in estimated coefficient. α_1 : 0.75

4.3 A More Extensive Investigation of the LDV Model

Here, we further explore the performance of the LDV model in the dynamic context. In this analysis, we add a variety of experimental conditions. First, we vary the sample size. Even under ideal conditions, OLS is only a consistent estimator, and given the small sample sizes often used in time series contexts, it is important to know the small sample properties of OLS with a lagged dependent variable. Second, we vary the autoregressive properties of the explanatory variable to assess how this affects the performance of the LDV model. And finally, we again pay attention to the autocorrelation in the error process of the dependent variable. This new set of experiments provides a much broader set of conditions to assess the robustness of the LDV model to violating the assumption that the error term is IID.

The DGP for Y remains that of equation (18), and the parameter values for the Y DGP remain: $\alpha_1 = 0.75$ and $\beta_1 = 0.50$. We vary the values for ρ_1 from 0.65 to 0.95 in increments of 0.10 to examine the effect of autocorrelation in the X DGP. We, again, set ρ_2 to three different values: 0.00, 0.10, and 0.20. Finally, we used sample sizes of 25, 50, 75, 100, 250, and 500.

Each Monte Carlo experiment was repeated 1000 times. We obtained a variety of information from the Monte Carlo experiments: the amount of bias in β_1 the parameter for the lone X variable, as well as both the rejection rate and overconfidence to assess the quality of the inferences.

4.4 Bias in the estimates of $\hat{\beta}_1$

Table 5 presents our Monte Carlo results for the bias in $\hat{\beta}_1$. $\hat{\beta}_1$ should be biased downwards, and this expectation is confirmed to some extent. But this is only true under certain circumstances, so the pattern of the bias does not exactly match the analytical expectations. While β_1 is generally underestimated when ρ_2 is 0.10 or above, this only occurs when the sample size is sufficiently large. The better estimates provided by larger samples appear to work at cross-purposes to the bias caused by the simultaneity. Regardless, the amount of bias is small even when the sample size is 25. The bias does increase as the autoregression in X strengthens, but, typically, the increase in bias between the lowest and highest amounts of autocorrelation in X is never more than 5%.

We see that OLS is asymptotically unbiased once the autocorrelation in the error process of Y is eliminated. The sample size need not be large to estimate the LDV model with OLS. Even when the sample size is 25 cases the bias is approximately 6% and falls to around three percent once N increases to 75.

4.5 Rejection Rates and Overconfidence for $\hat{\beta}_1$

Often the concern is that $\hat{\beta}_1$ will be too small, making it harder to confirm the effect of X on Y when it truly exists. In other words, it will be more difficult than it should be to reject the null hypothesis that $\hat{\beta}_1 = 0$. As such, we also present evidence from our Monte Carlo analysis regarding the ability of LDV models to generate proper causal inferences. Here, we ask two questions: how often is the null hypothesis for $\hat{\beta}_1$ correctly rejected, and how confident can we be in the OLS standard errors? The rejection rate is calculated as the percentage of times we fail to reject $\hat{\beta}_1 = 0$ out of the 1000 replications. The measure of overconfidence we use is one used by Beck and Katz (1995). With this measure, the quality of the OLS estimates of variability are assessed by calculating the ratio between the root mean square average of the one thousand estimated errors and the corresponding standard deviations of the one thousand estimates.[9] If the measure of overconfidence is above 100%, let's say 150%, then the true sampling variability is, on average, one and a half times the reported estimate of variability.

[9]The measure is more precisely: $100\frac{\sqrt{\sum^{1}000_{l=1}(\beta - b\bar{e}ta)^2}}{\sqrt{\sum^{1}000_{l=1}(s.e.(\hat{\beta}))^2}}$.

Table 5: Bias, As A Percentage, In $\hat{\beta}_1$, The Coefficient Of X

		$\rho_2 = 0.00$			
N	ρ_1	0.65	0.75	0.85	0.95
25		6.25	8.00	9.74	11.79
50		4.15	5.05	5.98	6.88
75		2.52	3.21	3.86	4.40
100		2.32	2.82	3.32	3.72
250		0.62	0.86	1.11	1.34
500		0.60	0.74	0.88	1.03
		$\rho_2 = 0.10$			
N	ρ_1	0.65	0.75	0.85	0.95
25		5.33	6.78	8.31	9.99
50		2.32	2.88	3.57	4.32
75		0.30	0.61	0.98	1.41
100		−0.09	0.02	0.24	0.54
250		−2.07	−2.25	−2.30	−2.18
500		−2.14	−2.43	−2.62	−2.60
		$\rho_2 = 0.20$			
N	ρ_1	0.65	0.75	0.85	0.95
25		4.44	5.60	6.88	8.19
50		0.50	0.68	1.06	1.61
75		−1.92	−2.04	−2.04	−1.80
100		−2.51	−2.88	−3.04	−2.95
250		−4.81	−5.51	−5.99	−6.12
500		−4.93	−5.75	−6.40	−6.67

Results are based on 1000 Monte Carlo replications.
Cell entries represent the average percentage
of bias in the estimated coefficient.

Table 6: Rejection Rate For $\hat{\beta}_1$, The Coefficient Of X

		$\rho_2 = 0.00$			
N	ρ_1	0.65	0.75	0.85	0.95
25		16.80	14.00	10.90	08.30
50		0.80	0.60	0.30	0.00
75		0.10	0.00	0.00	0.00
100		0.00	0.00	0.00	0.00
250		0.00	0.00	0.00	0.00
500		0.00	0.00	0.00	0.00
		$\rho_2 = 0.10$			
N	ρ_1	0.65	0.75	0.85	0.95
25		17.70	15.40	12.00	9.10
50		1.30	0.80	0.50	0.10
75		0.30	0.00	0.00	0.00
100		0.00	0.00	0.00	0.00
250		0.00	0.00	0.00	0.00
500		0.00	0.00	0.00	0.00
		$\rho_2 = 0.20$			
N	ρ_1	0.65	0.75	0.85	0.95
25		19.00	16.60	13.80	10.90
50		1.40	1.10	0.80	0.10
75		0.40	0.10	0.00	0.00
100		0.00	0.00	0.00	0.00
250		0.00	0.00	0.00	0.00
500		0.00	0.00	0.00	0.00

Results are based on 1000 Monte Carlo replications. Cell entries represent the percentage of time we fail to reject $\hat{\beta} = 0$.

Table 6 contains the percentage of times that we conclude X has no effect on Y despite the true effect being 0.50. The results show that this happens rarely even when the errors of Y are highly autocorrelated. Whatever the level of residual autocorrelation in the errors of Y, the rejection rate tends to be high only when the sample size is 25. For example, when $\rho_2 = 0.20$ and the sample size is 25, the null hypothesis will not be rejected 10-20% of the time. However, when the sample size increases to 50 cases, the rejection rate falls to 1%, indicating that making incorrect inferences is rare. Under the other conditions, once the sample size is 50 we incorrectly reject the null hypothesis less than 1% of the time. The evidence, here, emphasizes that, unless the N is extremely low, the bias is in most cases not enough to cause an analyst to make incorrect inferences.

Next, we turn to the measure of overconfidence. Again residual correlation causes few problems. The OLS standard errors are never more than 12% overconfident. And, once ρ_2 is 0.10, the OLS errors are rarely more than 5% overconfident. Only when the sample size is 25 are we overconfident by around 10%. The overconfidence shrinks even more when ρ_2 is 0.00. Once the sample size increases to 50 the OLS standard errors are never overconfident by more than two to three percent and is often less than that.

5 Comparing the Results

The results so far do not give an overall sense of how the performance of OLS changes as we move from the ideal condition of ρ_2 being 0.00 to when it rises to 0.10 or 0.20. We now compare the overall model performance across the three levels of ρ_2. To make the comparison, we hold ρ_1 at 0.85 and then plot the root mean square error (RMSE) across the six different sample sizes for each level of ρ_2.[10] We also plot the level of overconfidence for the same set of conditions. The results are in Figure 5.

The plot for the RMSE emphasizes how small the difference in overall model performance is when ρ_2 is either 0.10 or 0.20 as opposed to 0.00. The RMSE is practically indistinguishable. The difference between the level of overconfidence appears to be larger but is only one or two percent.

[10]The RMSE calculation, here, includes both model parameters: α_1 and β_1.

Table 7: Overconfidence For $\hat{\beta}_1$, The Coefficient Of X

		$\rho_2 = 0.00$			
N	ρ_1	0.65	0.75	0.85	0.95
25		106	105	108	106
50		98	98	97	97
75		98	98	98	97
100		100	100	100	101
250		99	99	98	98
500		100	100	101	102
		$\rho_2 = 0.10$			
N	ρ_1	0.65	0.75	0.85	0.95
25		109	109	110	111
50		102	102	102	101
75		102	102	103	102
100		104	105	105	106
250		103	104	104	103
500		104	105	106	106
		$\rho_2 = 0.20$			
N	ρ_1	0.65	0.75	0.85	0.95
25		113	114	116	116
50		106	107	107	107
75		106	107	108	108
100		109	110	111	111
250		109	110	110	108
500		109	111	112	112

Results are based on 1000 Monte Carlo replications. Cell entries represent the overconfidence in the estimates of $\hat{\beta}$. See text for details.

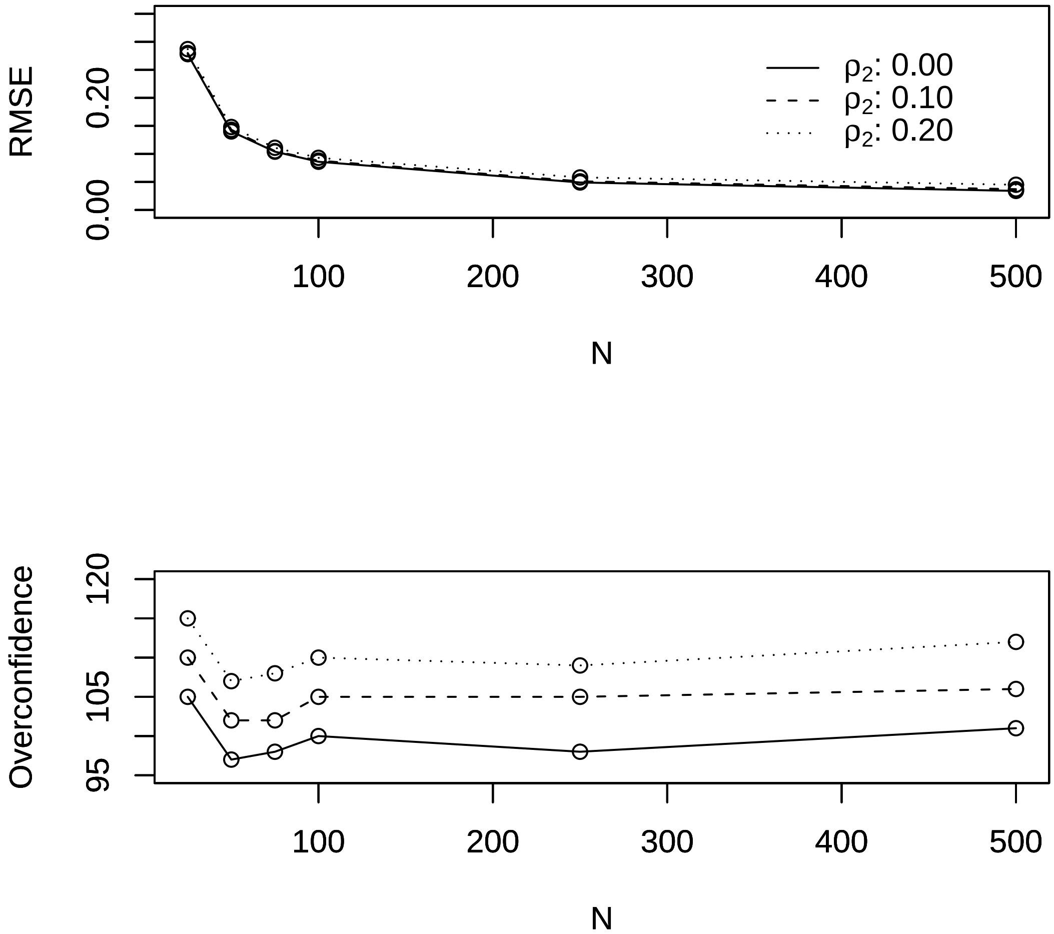

Figure 1: Comparisons of Model Performance

Figure 5 clearly emphasizes that for realistic levels of residual autocorrelation in Y, the performance of OLS is not greatly different than under conditions where OLS is asymptotically unbiased.

6 Remarks and Recommendations

Given the mountain of empirical results, we now distill what we have learned more generally about LDV models and develop some recommendations for the applied researcher who may be interested in whether he or she should use an LDV model. Here, we make four basic points.

First, researchers should be hesitant to use either GLS or OLS with corrected standard errors with autocorrelated data if one suspects the process is dynamic. Even when the process is only weakly dynamic, OLS was biased, and if the process is strongly dynamic, the bias caused by specification error in OLS and GLS was dramatic. The probability that a process is at least weakly dynamic is too great to ever use these models given the amount of bias found in the analysis.

Second, if the process is static, analysts should use ARMA models. When the DGP was static or weakly dynamic, the LDV model performed poorly while the ARMA model estimated with ML provided superior estimates.

Third, if the process is dynamic, however, the LDV provided estimates that were superior to the other models or estimators. But most importantly, the Monte Carlo evidence demonstrates that and LDV in the presence of residual autocorrelation does not induce significant amounts of bias. The RMSE across the differing values of ρ_2 was nearly identical. Moreover, the autoregressive nature of the X variable had little impact and large samples sizes were not required for good estimates. Even with as few as 50 cases, the LDV model estimates were quite good. Of course, after the estimation of any LDV model, the analyst should, of course, test that the model residuals are white noise through the use of a LaGrange Multiplier test. If the model residuals are still highly autocorrelated the LDV is probably inappropriate.

And finally, any analyst must test that the dependent variable is stationary before using an LDV model. Many of the problems that Achen encounters in the models he estimates as examples of where LDVs cause problems probably occur because the data are cointegrated. He readily admits that the budget data he considers are probably not stationary, in which case, the LDV model is the wrong model as techniques for cointegrated data should be used instead. Whatever the strengths of LDVs, they are inappropriate when the data are not stationary.

The real question, though, is how does one differentiate between the static and dynamic contexts? Here, the answers are less certain as there is no simple test for whether the data are static or dynamic. The analyst must ask whether the past matters to the current values of the process being studied? If the answer is yes, the LDV model is appropriate so long as the stationarity condition holds. The preponderance of the evidence in both economics and past work in political science is that most processes are dynamic. So, if one suspects that history matters, that the process has a memory, the LDV model is the best choice.

Appendix

A Stationarity Proof

We prove that the sum of α_1 and ρ_2 must be less than the absolute value of 1 for the model to be stationary. As usual we start with the following DGP:

$$\begin{aligned} y_t &= \alpha_1 y_{t-1} + \varepsilon \\ \varepsilon_t &= \rho_2 \varepsilon_{t-1} + v_t \end{aligned} \tag{20}$$

The addition of an explanatory variable makes no difference to the proof, so we leave it out for clarity. With the above equations, we can substitute in equivalent terms and rearrange to produce the following time series process:

$$\begin{aligned} y_t - \alpha_1 y_{t-1} &= \rho_2(y_t - \alpha_1 y_{t-1}) + v_t \\ y_t &= (\rho_2 + \alpha_1) y_{t-1} - \rho_2 \alpha_1 y_{t-2} + v_t \end{aligned} \tag{21}$$

The DGP is clearly an AR 2 process and the sum of $(\rho_2 + \alpha_1)$ and $-\rho_2\alpha_1$ must be less, in absolute value, than 1 for the process to be stationary given the stationarity conditions for an AR 2 process.

B The Effect of Model Fit

How well a model fits can also affect the overall performance of an estimator. To investigate how the model fit affected the performance of LDV models, we replicated the Monte Carlo analysis under two additional conditions. In the first condition, we set β to 0.10 and in the second to 1.50. The results from these analyses are below. In general, the results mirror those reported earlier. If the value of ρ_2 is 0.00 the estimates will be highly precise as the sample size increases regardless of how strong the effect of X is on Y. However, the size of the effect of X on Y is more critical when $\rho = 0.50$. Here, if the effect of X on Y is small, the bias can be as large as the true parameter value. However, if the effect is large the bias becomes trivial. Not surprisingly, strong effects are more impervious to the bias present in LDV models. Smaller effects are always harder to find and the bias here exacerbates this to some extent.

The overall fit of the models did not, except by sample size, vary much. Moreover, the overall model fit as summarized by the adjusted R^2 was not related to the amount of bias in the model. For the models where $\beta = 1.5$, the smallest average adjusted R^2 was .90 for 25 cases and was generally higher, even when the bias was fairly high under the $\rho_2 = 0.50$ condition. When β was .1 the adjusted R^2 often was quite low even when the bias was quite small. For example, when ρ_2 was 0.00 and the bias very small, the adjusted R^2 was as small as .38 for 25 cases and grew to .57 for 500 cases. So in general, the overall model fit was not indicative of how well OLS estimated the coefficients.

B.1 Results for β set to 0.10

Table 8: Bias in $\hat{\alpha}$, the coefficient of lagged Y

			$\rho_2 = 0.00$		
N	ρ_1	0.65	0.75	0.85	0.95
25		−22.01	−22.83	−23.62	−24.21
50		−10.71	−11.08	−11.51	−11.79
75		−7.18	−7.41	−7.62	−7.81
100		−4.89	−5.09	−5.37	−5.69
250		−2.09	−2.15	−2.22	−2.30
500		−1.05	−1.09	−1.16	−1.29
			$\rho_2 = 0.10$		
N	ρ_1	0.65	0.75	0.85	0.95
25		−15.02	−15.85	−16.72	−17.48
50		−4.42	−4.82	−5.32	−5.79
75		−1.20	−1.48	−1.80	−2.14
100		0.89	0.64	0.29	−0.17
250		3.40	3.27	3.11	2.87
500		4.32	4.21	4.05	3.76
			$\rho_2 = 0.20$		
N	ρ_1	0.65	0.75	0.85	0.95
25		−8.60	−9.43	−10.35	−11.26
50		1.22	0.81	0.27	−0.33
75		4.13	3.83	3.45	2.99
100		6.03	5.76	5.36	4.82
250		8.25	8.10	7.88	7.54
500		9.06	8.92	8.71	8.33

Results are based on 1000 Monte Carlo replications. Cell entries reprseent the average percentage of bias in the estimated coefficient.

Table 9: Bias in $\hat{\beta}$, the coefficient of X

			$\rho_2 = 0.00$		
N	ρ_1	0.65	0.75	0.85	0.95
25		9.58	14.64	21.57	34.30
50		6.91	10.00	14.68	21.95
75		3.39	6.18	9.71	14.59
100		3.89	5.54	8.04	11.54
250		−0.11	0.92	2.40	4.17
500		1.81	2.41	2.93	3.21
			$\rho_2 = 0.10$		
N	ρ_1	0.65	0.75	0.85	0.95
25		7.48	11.62	17.27	27.56
50		3.38	5.08	7.95	12.71
75		−0.71	0.56	2.11	4.44
100		−0.64	−0.53	−1.28	−2.88
250		−5.00	−5.58	−6.26	−7.50
500		−3.08	−4.11	−5.82	−8.75
			$\rho_2 = 0.20$		
N	ρ_1	0.65	0.75	0.85	0.95
25		5.53	8.90	13.42	21.48
50		0.34	0.78	1.98	4.34
75		−4.24	−4.34	−4.61	−4.65
100		−4.61	−5.91	−7.33	−9.21
250		−9.39	−11.40	−14.03	−18.06
500		−7.42	−9.91	−13.64	−19.54

Results are based on 1000 Monte Carlo replications. Cell entries represent the average percentage of bias in the estimated coefficient.

B.2 Results for β set to 1.50

Table 10: Bias in $\hat{\alpha}$, the coefficient of lagged Y

		$\rho_2 = 0.00$			
N	ρ_1	0.65	0.75	0.85	0.95
25		−2.92	−2.48	−2.08	−1.75
50		−1.23	−0.94	−0.70	−0.51
75		−0.77	−0.61	−0.48	−0.32
100		−0.57	−0.45	−0.34	−0.25
250		−0.24	−0.17	−0.12	−0.09
500		−0.05	−0.05	−0.06	−0.08
		$\rho_2 = 0.10$			
N	ρ_1	0.65	0.75	0.85	0.95
25		−2.07	−1.79	−1.55	−1.38
50		−0.54	−0.42	−0.31	−0.24
75		−0.09	−0.10	−0.11	−0.07
100		0.09	0.05	0.04	−0.11
250		0.41	0.32	0.22	0.13
500		0.61	0.44	0.28	0.13
		$\rho_2 = 0.20$			
N	ρ_1	0.65	0.75	0.85	0.95
25		−1.15	−1.05	−0.99	−0.99
50		0.24	0.18	0.14	0.07
75		0.69	0.50	0.32	0.23
100		0.88	0.64	0.44	0.27
250		1.18	0.90	0.65	0.41
500		1.39	1.04	0.71	0.41

Results are based on 1000 Monte Carlo replications. Cell entries represent the average percentage of bias in the estimated coefficient.

Table 11: Bias in $\hat{\beta}$, the coefficient of X

		$\rho_2 = 0.00$			
N	ρ_1	0.65	0.75	0.85	0.95
25		1.49	1.75	1.95	2.09
50		0.81	0.87	0.93	1.01
75		0.42	0.51	0.57	0.59
100		0.54	0.58	0.59	0.58
250		0.09	0.11	0.12	0.15
500		0.10	0.13	0.17	0.21
		$\rho_2 = 0.10$			
N	ρ_1	0.65	0.75	0.85	0.95
25		1.38	1.60	1.79	1.88
50		0.47	0.51	0.56	0.66
75		−0.04	0.030	0.10	0.15
100		0.05	0.07	0.10	0.13
250		−0.48	−0.48	−0.44	−0.37
500		−0.49	−0.48	−0.42	−0.33
		$\rho_2 = 0.20$			
N	ρ_1	0.65	0.75	0.85	0.95
25		1.27	1.44	1.61	1.67
50		0.09	0.10	0.15	0.26
75		−0.56	−0.52	−0.44	−0.35
100		−0.51	−0.52	−0.48	−0.41
250		−1.16	−1.18	−1.13	−1.00
500		−1.20	−1.21	−1.13	−0.97

Results are based on 1000 Monte Carlo replications. Cell entries represent the average percentage of bias in the estimated coefficient.

References

Achen, Christopher H. 2000. "Why Lagged Dependent Variables Can Supress the Explanatory Power of Other Independent Variables." Presented at the Annual Meeting of the Political Methodology, Los Angeles.

Bartels, Larry M. 1991. "Instrumental and 'Quasi-Instrumental' Variables." *American Journal of Political Science* 35:777–880.

Beck, Nathaniel. 1985. "Estimating Dynamic Models is not Merely a Matter of Technique." *Political Methodology* 11:71–89.

Beck, Nathaniel. 1992. "Comparing Dynamic Specifications: The Case of Presidential Approval." *Political Analysis* 3:27–50.

Beck, Nathaniel and Jonathan N. Katz. 1995. "What to Do (And Not to Do) With Time-Series Cross-Section Data." *American Political Science Review* 89:634–647.

Davidson, Russell and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics.* New York: Oxford University Press.

Griliches, Zvi. 1961. "A Note of Serial Correlation Bias in Estimates of Distributed Lags." *Econometrica* 29:65–73.

Hendry, David F. 1995. *Dynamic Econometrics.* Oxford: Oxford University Press.

Hendry, David and Grayham Mizon. 1978. "Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England." *Economic Journal* 88:549–563.

Hibbs, Douglas A., Jr. 1974. Problems of Statistical Estimation and Causal Inference in Time-Series Regression Models. In *Sociological Methodology 1973-1974*, ed. Herbert L. Costner. San Francisco: Jossey-Bass pp. 252–308.

Hurwicz, L. 1950. Least-Squares Bias in Time Series. In *Statistical Inference in Dynamic Economic Models*, ed. T. Koopmans. New York: Wiley pp. 215–249.

Maddala, G.S. and A.S. Rao. 1973. "Tests for Serial Correlation in Regression Models with Lagged Dependent Variables and Serially Correlated Errors." *Econometrica* 47:761–774.

Malinvaud, E. 1970. *Statistical Methods of Econometrics.* 2nd ed. Amsterdam: North-Holland.

Mizon, Grayham. 1977. "Inferential Procedures in Nonlinear Models: An Application In A UK Industrial Cross-Section Study of Factor Substitution In Returns to Scale." *Econometrica* 45:1221–1242.

Mizon, Grayham. 1995. "A Simple Message for Autocorrelation Correctors - Don't." *Journal of Econometrics* 69:267–288.

Phillips, P.C.B. 1977. "Approximations To Some Finite Sample Distributions Associated With A First-Order Stochatic Difference Equation." *Econometrica* 45:463–486.

Phillips, P.C.B. and M.R. Wickens. 1978. *Exercises in Econometrics.* Vol. 2 Oxford: Phillip Allan.

White, J. 1961. "Asymptotic Expansions For The Mean and Variance Of The Serial Correlation Coefficient." *Biometrika* 48:85.–94.