Political Analysis Advance Access published November 23, 2005

doi:10.1093/pan/mpj006

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

Luke Keele
Department of Political Science, Ohio State University,
154 N. Oval Mall, Columbus, OH 43210
e-mail: luke.keele@mail.polisci.ohio-state.edu (corresponding author)

Nathan J. Kelly
Department of Political Science, University of Tennessee,
1001 McClung Tower, Knoxville, TN 37996-0410
e-mail: nathan.j.kelly@gmail.com

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 assess empirically 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 an appropriate model for the dynamic theories often tested by applied analysts. From the analysis, we develop several practical suggestions on when and how to 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 estimator 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 estimator assumes, for example, that there is no autocorrelation in the residuals. That is, the residual at one point of observation is not correlated with any other residual. In time series data, of course, this

Authors' note: A previous version of this article was presented at the 2003 Southern Political Science Meeting. For suggestions and criticisms, we thank Jim Stimson, Neal Beck, Chris Wlezien, Chris Zorn, Bob Andersen, Mike Colaresi, and four anonymous reviewers. Replication materials and an extra appendix are available on the *Political Analysis* Web site.

assumption is almost always violated. One common view of autocorrelation is that it is a technical violation of an OLS assumption that leads to incorrect estimates of the standard errors. A second view of autocorrelation involves thinking of time series data in the context of political dynamics. Instead of mechanistically worrying about the residuals, we can develop theories and use specifications that capture the dynamic processes in question. From this perspective, analysts view autocorrelation as a potential sign of improper theoretical specification rather than a technical violation of an estimator assumption (Hendry and Mizon 1978; Beck 1985; Mizon 1995).

Regardless of which perspective on autocorrelation is adopted, lagged dependent variables have been proposed and utilized (for good or ill) as a solution. When one detects autocorrelation in an OLS regression, the inclusion of a lagged dependent variable often eliminates any residual serial correlation. This is an example of using a lagged dependent variable under the view of autocorrelation simply as a violation of an estimator assumption. Lagged dependent variables are also utilized as a means of capturing the dynamics of politics. In the study of public opinion, for example, there are theories in which an attitude at time t is a function of that same attitude at $t - 1$ as modified by new information. Here, the lagged dependent variable captures a theory of dynamics with a dynamic specification.

The extant literature, however, presents a case against using lagged dependent variables with OLS, with Achen (2000) making the most recent presentation of the argument. Achen proves (and we will revisit this well-known finding below) that inserting a lagged dependent variable is a dangerous strategy for ridding the residuals of autocorrelation because coefficient estimates can be biased. Achen also applies this argument to the second view of autocorrelation. Even when a lagged dependent variable is theoretically appropriate, remaining residual autocorrelation can lead to biased coefficient estimates. It would appear, then, that we are faced with the following quandary: On the one hand, including a lagged dependent variable can lead to biased coefficient estimates. On the other hand, excluding a theoretically appropriate lagged dependent variable produces misspecification. Our goal is to resolve this quandary.

Given the ease of estimating these models with OLS and the theoretical power of lagged dependent variables, it remains useful to know when and if they can be used. The question then becomes, is it ever appropriate to use OLS to estimate a model with a lagged dependent variable? The dominant response to this question in our discipline used to be yes. Lagged dependent variable models were once estimated with great frequency. If there was a problem with autocorrelation or if an analyst merely wanted to control for some unspecified spurious correlation, the lagged dependent variable model was viewed as a completely reasonable corrective procedure. More recently, however, we have come almost full circle. After learning about the biases that the introduction of lagged dependent variables can produce, they have come to be viewed with great scepticism. It seems that the current default response to lagged dependent variables is suspicion if not outright objection.

The research we report in this article leads us to the conclusion that lagged dependent variable models estimated with OLS are appropriate under certain conditions and inappropriate under others. Our primary goal is to identify these conditions and compare the performance of lagged dependent variable models to common alternatives. We begin with a discussion of the properties of OLS with a lagged dependent variable, and we define precisely the conditions under which problems may arise. The analytic results that we discuss may be well known to some readers, but they provide the starting point for the rest of our discussion. We then conduct a series of Monte Carlo experiments that allow for an examination of the properties of lagged dependent variables (and their alternatives) in the context of the small samples that are so common in time series analyses. The Monte Carlo analysis provides us with the ability to identify the applied conditions under which a lagged dependent variable is most and least appropriate. We conclude with a discussion of our results and some general guidelines regarding when a lagged dependent variable should be included on the right-hand side.

2 The Logic and Properties of Lagged Dependent Variables

In this section, we begin with a conceptual discussion of lagged dependent variables (LDV). 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 an LDV, 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 OLS when used with 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:

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

Specifically, Eq. (1) 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 $\beta = 0$, we are left with a lagged dependent variable model:

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

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

Earlier, we discussed two reasons for estimating such a model. The first is to rid the residuals of serial correlation. While in practice the inclusion of a lagged dependent variable will accomplish this, a theoretically motivated reason to include a lagged dependent variable is to capture, through specification, a type of dynamics that frequently occurs in politics. We use evaluations of the president as an illustrative 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 estimate the following model:

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

[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.

The lagged X's will undoubtedly be highly collinear, leading to imprecise estimates of the β's. A better way to proceed would be to assume some functional form for how the 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. To do this, we add a multiplier, λ, to the equation to be estimated, which induces geometric decay in the effect of X_t on Y_t. Using some algebraic manipulation, we can capture this multiplier effect with the coefficient on lagged Y_t:

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

Therefore, 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 not only resonate in the current quarter but will 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 coefficient has a dynamic interpretation as it dictates the timing of the effect of X_t on Y_t. This makes it a good choice for situations in which theory predicts that the effects of X_t 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 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 LDVs 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 data-generating process (DGP), or true underlying relationship (Davidson and

[3]The steps required to go from the second to the third part of Eq. (4) are not entirely trivial in that the mathematics raise an important issue about the error term. The last line of Eq. (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 nontrivial part of this equation is the error term, $u_t - \lambda u_{t-1}$, which is an 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.

MacKinnon 1993). The proof of this appears infrequently, so we reproduce it to help clarify the issues that surround the use of OLS to estimate LDV models. Consider a simple example where:

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

We assume that $|\alpha| < 1$ 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 Eq. (5) into Eq. (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 Eq. (8) is, at this point, 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 DGP has what we call a common factor term as opposed to a dynamic term (that is, there are no dynamics in the form of lagged values of Y_t on the right-hand side).[5] To understand

[4]That is not to say they are impossible to derive. Hurwicz (1950), White (1961), and Phillips (1977) have all derived the small sample properties of α analytically but only for the case in which ϕ is 0.0.

[5]The common factor we refer to is $(1 - \beta_2 L)$. See Hendry (1995) for a more in-depth treatment.

what we mean by a common factor DGP, we write the DGP with the following equations:

$$Y_t = \alpha Y_{t-1} + \beta X_t + u_t \tag{10}$$

where:

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

and

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

For the DGP above, α is the dynamic parameter, and ρ and ϕ are the autoregressive parameters for X_t and the error term of Y_t, respectively. This implies that X_t and u_t are equal to their value in the previous period times ρ and ϕ plus some new component e_{1t} and e_{2t}, respectively. In this discussion, we assume that all three processes are stationary, that is, α, ρ, and ϕ are all less than one in absolute value. We are in the common factor context when $\alpha = 0$ and the autoregressive properties of the model are due solely to autocorrelation in the error term. Achen considers the consequences of incorrectly including a lag of Y_t with a common factor DGP. He demonstrates that the estimates of both α and β will be biased by the following amounts if an LDV is incorrectly included in the estimating equation:

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

$$\text{plim } \hat{\alpha} = \alpha + \phi\frac{1 - R^2}{1 - \rho^2 R^2}. \tag{14}$$

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_t). Clearly, the estimate of β will be downwardly biased as both ϕ and ρ increase. In this 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 in which 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_t, produces bias instead of inefficiency. 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 LDVs, however, is more far reaching.

Achen next considers the case in which the lag of Y_t is correctly part of the DGP, a situation in which the DGP is dynamic or more precisely when $\alpha > 0$. He derives the asymptotic bias in α and β for the dynamic context:

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

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

where $s^2 = \sigma^2_{Y_{t-1},X_t}$ and $g = \text{plim}(\hat{\alpha} - \alpha)$. If ϕ is zero, then OLS with an LDV is consistent, as demonstrated earlier. However, if $\phi \neq 0$, the estimate of β will be biased downward as ρ and α increase (Griliches 1961; Malinvaud 1970; Maddala and Rao 1973; Hibbs 1974;

Phillips and Wickens 1978; Achen 2000). It would appear, then, that if the errors are autocorrelated at all, even in the dynamic context, fitting a model that has an LDV with OLS is problematic and inadvisable.

Before the last rites are read, however, four factors suggest that LDV models fitted with OLS might be appropriate under certain conditions.

- First, for Eq. (10) to be stationary, the following condition must be satisfied: $|\alpha + \phi| < 1$.[6] When this stationarity condition is satisfied, residual autocorrelation (ϕ) is limited by a high value for α. The stationarity condition is not trivial, because if it is violated we have a permanent memory process and a different set of statistical techniques apply. This limit on the value of ϕ implied by the stationarity condition suggests that there will be decreasing amounts of autocorrelation in the residuals as the DGP becomes more dynamic, and therefore places a limit on the bias.
- Second, while OLS without an LDV utilizing Newey-West standard errors may be best when α is 0, if α is nonzero, OLS without an LDV will be biased due to an omitted variable. In short, if $\alpha \neq 0$, omitting the lag of Y_t 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 DGP 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 even if the residuals are autocorrelated.
- Third, the analytical results in the dynamic context must rely on values of s^2, the covariance between Y_{t-1} and X_t, that can only be set arbitrarily. Furthermore, the analytic results discussed above apply to infinitely large samples, and time series analysts rarely deal with such large t's. While we can pick a range of values for s^2 and the size of the sample, in reality we need experimental evidence in which both of these values vary as they might with real data.
- Finally, the estimation of an LDV model with OLS is problematic only when residual autocorrelation is present. As in other time series models, if the estimated residuals show significant amounts of autocorrelation, then either a different specification or a different estimator is needed. The real question is whether tests for residual serial correlation are sensitive enough to prevent an analyst from reporting biased OLS estimates.

In short, in the real world of hard choices, it matters whether we are facing a tiny expected bias or a larger value that fundamentally alters the magnitude of the coefficients. The analytic results imply bias, but they 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 analytic results are more difficult to derive.

What follows, then, is a series of Monte Carlo experiments. We begin by asking a very basic question: Are LDV models ever an appropriate choice for either the common factor or dynamic situations defined above?

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

3 Monte Carlo Analysis

In the Monte Carlo analysis, we use the same general DGP from above:

$$Y_t = \alpha Y_{t-1} + \beta X_t + u_t \tag{17}$$

$$X_t = \rho X_{t-1} + e_{1t} \tag{18}$$

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

where $e_{1t}, e_{2t} \sim IID(0, \sigma^2)$. The DGP allows for both the regressor and the error term to be autoregressive; however, we restrict both ρ and ϕ to be less than one in absolute value, making them both stationary. We vary the values of α, ρ, and ϕ as our experimental manipulations, but with a focus on nonzero ϕ given that α is either small or large.

In all of the experiments reported below, we examine bias in the estimates of β, which we report as a percentage—that is, the percentage the average estimate is above (positive values) or below (negative values) the true parameter value. There are two biases that we can report. The first is simply $E(\hat{\beta}) - \beta$. But when α is greater than zero, the total effect of X_t on Y_t is no longer β but is instead $\frac{\beta}{1-\alpha}$, as noted earlier. We also then report the average difference between the estimated total effect and the true total effect.

We compare the performance of the LDV model estimated with OLS to a variety of alternative procedures. The first of these alternative estimators is an ARMA(1, 0) model estimated via MLE. Second, we include two models fitted with GLS. Specifically, we estimate Cochrane-Orcutt and Prais-Winsten regressions. Both Cochrane and Prais assume AR(1) error structures but no dynamics. Cochrane-Orcutt is asymptotically equivalent to the use of nonlinear least squares to estimate an AR(1) model, while Prais-Winsten is equivalent to the use of full information ML for an AR(1) model. We use OLS to estimate a model with a second lag of Y_t included on the right-hand side. Finally, we use OLS to estimate a model without any lags of Y_t on the right-hand side.[7] These estimators and models were chosen as being the most likely alternatives for dealing with autocorrelation. Please see the on-line documentation for further details on these estimators and models.

Based on the analytical results discussed previously, the performance of OLS with an LDV compared to OLS without an LDV should move in opposite directions in response to changes in α. As the DGP becomes more dynamic (as the value of α 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 increases. These gains in performance, however, may be minimal if the errors are autocorrelated. The expectations for the other estimators and models are less clear-cut.

We conduct two rounds of experiments. In the first round, we conduct two experiments designed to address the basic question of whether the estimation of an LDV model can ever be appropriate. The fundamental issue is how best to estimate the effect of X_t on Y_t. In the second round, we explore the implications of the first set of experiments in more detail.

3.1 *Are LDV Models Ever Appropriate?*

In the first experiment, we fix ϕ to 0.75 and study situations in which an LDV model is fitted to a DGP with a common factor, $\alpha = 0.0$, to what we will call a weakly dynamic DGP

[7]We include OLS, since it should be unbiased so long as α is 0.0. The use of OLS without lags in the model would require an analyst to use Newey-West standard errors, which we do not calculate because we are concerned only with bias.

Table 1 Comparative percentage of bias for common factor to weakly dynamic DGP

	α			
Model	*0.00*	*0.10*	*0.20*	*0.50*
LDV	−55.25	−55.68	−55.57	−50.68
ARMA	0.01	2.88	4.61	3.41
GLS-Cochrane	0.49	4.17	7.07	12.67
OLS[a]	0.36	10.23	22.27	81.68
GLS-Prais	0.08	3.82	6.83	12.87
2LDV	−55.03	−54.84	−54.60	−53.49

Note. Results are based on 1000 Monte Carlo replications. Cell entries represent the percentage of bias in β, the average estimated coefficient of X_t. ϕ: 0.75; ρ: 0.95.
[a]Does not include any lags.

(α = 0.10, 0.20, or 0.50). This set of experiments mimics the first situation outlined by Achen, one in which the LDV is not part of the DGP and either does not belong in the model or is a relatively minor component of the true process. In the second experiment, we create a dynamic DGP. Here, we fix α to 0.75, while ϕ is set to 0.00, 0.10, 0.20, or 0.50. This set of experiments examines situations ranging from those in which OLS with an LDV is consistent (no residual autocorrelation) to situations in which we have violated the assumption of IID errors to varying degrees. In both experiments, β is set to 0.50, t is fixed at 100, and each Monte Carlo experiment was repeated 1000 times. We also set ρ to 0.95 to make the test against the LDV as stringent as possible. As can be seen in the previous section, higher amounts of autoregression in X_t also has an adverse impact on the OLS estimates of the LDV model.

Experiment 1: The Common Factor to Weakly Dynamic DGP. We first report the results for the common factor to weakly dynamic DGP. We examine the accuracy of the estimated effect of X_t on Y_t. Tables 1 and 2 report the bias in the estimates of β as percentage. That is, each entry is the percentage of the bias in the average coefficient. For example, an entry of −5 means that the average estimate is 5% smaller than the true value of 0.50. If the entry is 5, the average estimate is 5% bigger than the true value of 0.50. Table 1 is for the initial effect of X_t on Y_t, while Table 2 reports the bias in the total effect of X_t on Y_t.

Table 2 Comparative bias in long-run multiplier effect for common factor to weakly dynamic DGP

	α			
Estimator	*0.00*	*0.10*	*0.20*	*0.50*
LDV	18.83	21.86	25.26	38.37
ARMA	0.01	−7.40	−16.30	−48.30
GLS-Cochrane	0.49	−6.25	−14.33	−43.67
OLS[a]	0.36	−0.78	−2.18	−9.16
GLS-Prais	0.08	−6.55	−14.53	−43.56
2LDV	46.58	130.19	242.09	−301.07

Note. Results are based on 1000 Monte Carlo replications. Cell entries represent the percentage of bias in β, the average estimated coefficient of X_t. Long-run effect: $\frac{\beta}{1-\alpha}$; ϕ: 0.75; ρ: 0.95.
[a]Does not include any lags.

Table 3 Comparative bias for dynamic DGP

	ϕ			
Model	*0.00*	*0.10*	*0.20*	*0.50*
LDV	3.91	0.70	−2.83	−17.17
ARMA	−0.95	−1.59	−2.14	−3.81
GLS-Cochrane	74.72	68.27	61.46	37.41
OLS[a]	201.48	201.35	201.19	200.52
GLS-Prais	76.59	70.16	63.38	39.25
2LDV	4.05	−1.67	−7.66	−28.51

Note. Results are based on 1000 Monte Carlo replications. Cell entries represent the percentage of bias in β, the average estimated coefficient of X_t. α: 0.75; ρ: 0.95.
[a]Does not include any lags.

Under the common factor DGP, the bias is minimal except when one or more lags of Y_t are included on the right-hand side. Here, the bias is considerable. If an LDV is wrongly included when the DGP contains a common factor, β will be underestimated by around 55%, while the bias for the other estimators is around 2–3%. The bias tends to increase across the other estimators as the magnitude of α grows. The effect of the omitted variable bias is particularly noticeable when OLS is used without any lags in the model. The bias is less than 1% for the common factor DGP but is a sizeable 81% once α is 0.50. The bias here is in the opposite direction from that of OLS with an LDV, as $\hat{\beta}$ is too large.

Under a weakly dynamic DGP, the bias for the total effect of X_t on Y_t grows dramatically as α increases for most of the estimators. This is not surprising since, with the exception of the LDV model, these setups do not provide an estimate of α and, given how the total effect is defined, $\frac{\beta}{(1-\alpha)}$, an accurate initial estimate of β will be increasingly inaccurate as the size of the true α grows. Interestingly, OLS without any lags on the right-hand side proves to be an exception. Here, the omitted variable bias from leaving the lag out inflates the initial estimate of β, making it closer to the true *total* effect. We also see that the direction of bias in the LDV estimate of the total effect of X_t on Y_t reverses as compared to the partial effect captured in $\hat{\beta}$. This is because the total estimated effect is produced from both $\hat{\alpha}$ and $\hat{\beta}$. Positive bias in $\hat{\alpha}$ more than cancels out the negative bias in $\hat{\beta}$.

The consequences of fitting a dynamic model in the form of an LDV to a common factor DGP are clear. However, we should not be surprised that fitting the wrong model to the DGP produces biased results. So on this score, Achen is correct: an LDV with OLS is clearly inadvisable under these conditions. That said, it should be noted that except with the common factor DGP, none of the estimators do particularly well when the DGP contains both a common factor and dynamics. We next perform the same experiment for a dynamic DGP.

Experiment Two: The Moderate to Strongly Dynamic DGP. To simulate a dynamic DGP in this second experiment, we set α to 0.75, while ϕ is 0.0, 0.1, 0.20, or 0.50. The critical test now is how the performance of OLS with an LDV changes when ϕ is above 0.0.

Tables 3 and 4 contain the bias in $\hat{\beta}$, as a percentage, for all the estimators in the dynamic context. The bias for OLS with an LDV is a minimal 4% when there is no residual autocorrelation. The bias in the LDV model drops and then increases (in the negative direction) as the autoregressive nature of the error term increases. Only the ARMA model produces similar amounts of bias, at least for the initial effect of X_t on Y_t. In comparing the

Table 4 Comparative bias in the long-run multiplier effect for dynamic DGP

	ϕ			
Model	*0.00*	*0.10*	*0.20*	*0.50*
LDV	−1.12	0.58	2.69	14.73
ARMA	−75.23	−75.40	−75.54	−75.95
GLS-Cochrane	−56.32	−57.93	−59.63	−65.65
OLS[a]	−24.63	−24.66	−24.70	−24.87
GLS-Prais	−55.85	−57.46	−59.16	−65.19
2LDV	21.09	34.36	49.46	−202.47

Note. Results are based on 1000 Monte Carlo replications. Cell entries represent the percentage of bias in β, the average estimated coefficient of X_t. Long-run effect: $\frac{\beta}{1-\alpha}$; α: 0.75; ρ: 0.95.
[a]Does not include any lags.

bias for total effect, however, the performance of the ARMA model is clearly worse than OLS with an LDV. What is most noticeable are the extremely poor estimates from GLS and OLS without any lags. For GLS, the estimates are nearly 70% too large, while for OLS without a lag, the estimates are over 200% too large. Clearly, both OLS without an LDV and GLS are poor choices when the data-generating process is dynamic.

From the analysis thus far, it would appear that the ARMA model and OLS with a single LDV produce the best estimates under the widest number of experimental conditions. To further compare the performance of these two estimators, we plot the bias (no longer as a percentage) in the total effect of X_t on Y_t as a function of a range of values for both α and ϕ. Plotting the bias in the total effect of X_t on Y_t is particularly instructive, since it captures the bias in both β and α. For the total effect, if α is underestimated, so too is the total effect, but the degree of underestimation is partially dictated by the estimate of β. If the estimate of α is accurate, the bias in the total effect is solely a function of the estimate of β. And if α is overestimated, the total effect will be too large, but the amount of overestimation again partially depends on the estimate of β. The ARMA model provides no estimate of α, so it will always underestimate the true total effect except in cases in which the true value of α is zero.

In Fig. 1 we see that the pattern for OLS is fairly complex. Here, the bias is a function of both α and ϕ. So long as both α and ϕ are not large, the estimates are good, but the estimates become quite bad for large values of both parameters. Figure 2 presents the same information for the ARMA model. The pattern in the bias here is obvious. As Fig. 2 makes clear, the bias in the total effect is solely a function of the value of α and can be quite large. The estimate of the total effect is over 150% too large when α is larger than 0.70.[8]

The statistical evidence suggests that OLS with an LDV is robust to modest violations of the assumption that the error term is IID, but if the error term is strongly autoregressive, the bias can be quite large. In the next section, we explore the implications of our findings in two ways. First, we further explore the performance of OLS with an LDV under a wider variety of dynamic contexts to see if the bias remains minor. Second, we test our ability to detect the autoregressive error that should signal when bias is present.

3.2 *A More Extensive Investigation of the LDVs in the Dynamic Context*

In the second major component of our analysis, we add a variety of experimental conditions to further explore the performance of OLS with an LDV in the dynamic context.

[8]The reader should note that for some of the values of α and ϕ in the plot, the model is no longer stationary. We had to do this to make the surface rectangular.

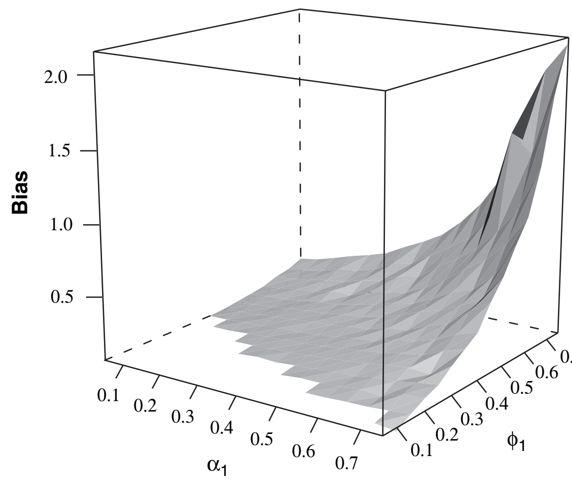

Fig. 1 Bias in the long-run effect, across α and φ for the LDV model.

First, we vary the sample size. Even under ideal conditions, OLS produces consistent estimates for the LDV model, 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 as an additional assessment of the LDV model performance. As before, we vary the autocorrelation in the error process to further understand how the bias changes as φ grows. Finally, we see how well a common test for residual autocorrelation detects the problematic autoregressive model error. This new set of experiments provides a much broader set of conditions for assessing the robustness of OLS with an LDV.

The DGP for Y_t remains that of Eq. (17), and the parameter values for the DGP remain: $\alpha = 0.75$ and $\beta = 0.50$. We vary the values for ρ from 0.65 to 0.95 in increments of 0.10 to examine the effect of autocorrelation in the X_t DGP. We again set φ to four different values: 0.00, 0.10, 0.20, and 0.50. Finally, we use sample sizes of 25, 50, 75, 100, 250, 500, and 1000. Each Monte Carlo experiment was repeated 1000 times.

We obtained a variety of information from the Monte Carlo experiments. We focus on the amount of bias in β, the parameter for the X_t variable, the rejection rate of the hypothesis that X_t has no effect on Y_t, and the results from a test for serial correlation in

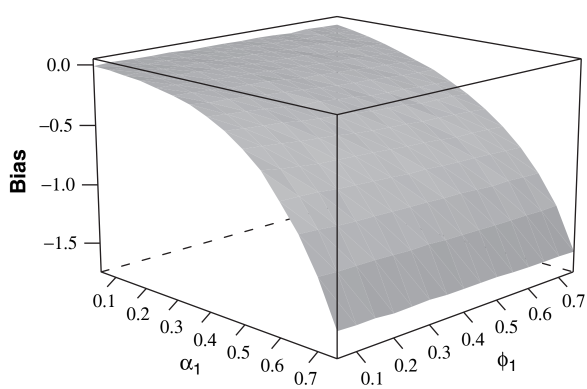

Fig. 2 Bias in the total effect, across α and φ for the ARMA model.

Table 5 Asymptotic least squares bias (as a percentage) for β, the coefficient of X_t

	ϕ			
ρ	*0.00*	*0.10*	*0.20*	*0.50*
0.65	0.00	−1.37	−2.98	−10.15
0.75	0.00	−1.85	−4.03	−13.71
0.85	0.00	−2.53	−5.52	−18.76
0.95	0.00	−3.57	−7.77	−26.43

the estimated residuals. We used the Breusch-Godfrey Lagrange Multiplier (LM) test for autocorrelation and calculated the percentage of times we rejected the null hypothesis of no serial correlation.

3.3 *Bias in the estimates of* $\hat{\beta}$

To provide a standard for the amount of bias we observe, we calculated the asymptotic bias for the LDV model when estimated with OLS for the experimental conditions used in the analysis. Table 5 provides the asymptotic bias calculations, again reported as a percentage. Even here we see that when ϕ is 0.10 the bias is not large but grows dramatically as ϕ increases. The question, however, is how much larger the bias will be with small samples.

Table 6 presents our Monte Carlo results for the bias in $\hat{\beta}$. We see that OLS is asymptotically unbiased when the autocorrelation in the error process of Y_t is not present. Here, even when the sample size is 25 cases, the bias is approximately 9% and falls to around less than 3% once t increases to 75. Next, we induce serial correlation in the errors. Now, $\hat{\beta}$ should be biased downward, and this expectation is confirmed to some extent. But the underestimation of $\hat{\beta}$ is true only under certain circumstances, so the observed pattern of bias does not match the analytical expectations. While β is generally underestimated when ϕ is 0.10 or above, this occurs only when the sample size is sufficiently large. Moreover, the bias tends to grow with t and tends to be less than the calculated asymptotic bias with the best estimates of β occurring when t is between 75–125—an unusual situation.

Further investigation revealed that this pattern in the bias is due to how OLS estimates $\hat{\alpha}$ when ϕ is greater than 0.0. We found a clear pattern in the bias of $\hat{\alpha}$, but it runs contrary to what one might expect. Table 7 contains the bias in $\hat{\alpha}$ when ϕ is 0.10 but there is no other regressor in the model. For small t, $\hat{\alpha}$ is underestimated, but that bias changes to an overestimation as t increases. The asymptotic bias under these conditions should be about 0.04. We find that the difference between the asymptotic bias and the simulated bias with 20000 cases is 9.3×10^{-3}. So with a large enough t, the estimated bias will converge to the asymptotic value. What does this imply for the estimates of β? Recall that the formula for the bias in β is:

$$\text{plim } \hat{\beta} = \left[1 - \frac{\rho g}{1 - \rho\alpha}\right]\beta. \tag{20}$$

This formula clearly implies that β is overestimated when $\hat{\alpha}$ is too small and β is underestimated when $\hat{\alpha}$ is too large. Since the estimates of α here vary systematically by sample size, this means that for small sample sizes $\hat{\beta}$ will be too large and for large sample sizes $\hat{\beta}$ will be too small. But in sample sizes for which $\hat{\alpha}$ is highly precise, around

Table 6 Bias, as a percentage, in $\hat{\beta}$, the coefficient of X

t	ρ	*0.65*	*0.75*	*0.85*	*0.95*
		$\phi = 0.00$			
25		8.98	10.85	12.52	11.93
50		3.77	4.92	5.60	8.11
75		2.38	3.44	3.76	5.31
100		2.08	2.92	2.81	2.92
250		1.55	1.27	1.13	1.37
500		0.83	1.08	0.75	0.77
1000		0.22	0.19	0.31	0.55
		$\phi = 0.10$			
25		5.65	6.01	7.52	8.18
50		2.44	2.41	3.88	3.89
75		0.63	0.08	2.01	2.01
100		−0.24	−1.39	0.09	0.16
250		−1.62	−1.49	−2.09	−2.11
500		−2.17	−2.41	−2.72	−2.90
1000		−2.55	−3.16	−3.32	−3.26
		$\phi = 0.20$			
25		5.76	5.92	8.52	7.59
50		0.71	1.74	1.14	1.19
75		−0.80	−0.90	−2.22	−1.03
100		−3.16	−3.05	−3.96	−2.86
250		−4.26	−4.95	−5.72	−6.17
500		−5.31	−5.65	−6.53	−6.87
1000		−5.51	−6.11	−7.23	−7.44
		$\phi = 0.50$			
25		−1.30	3.50	6.17	−0.79
50		−4.85	−7.09	−9.80	−8.58
75		−8.83	−11.02	−13.38	−14.07
100		−9.45	−12.90	−15.29	−16.45
250		−13.77	−15.89	−19.94	−21.74
500		−14.31	−17.61	−21.22	−23.94
1000		−15.24	−18.48	−21.86	−25.03

Note. Results are based on 1000 Monte Carlo replications. Cell entries represent the average percentage of bias in $\hat{\beta}$, the estimated coefficient of X_t.

75–125, the estimates of β will be very accurate. It is this surprising pattern in the estimates of α that is responsible for the pattern in the bias in $\hat{\beta}$, whereby we see a positive bias for small t that converges to zero before reversing into a negative bias for larger sample sizes that grows until it converges to the asymptotic values. This is a counterintuitive result that makes the bias for LDV models quite small for typical sample sizes. Please see the appendix for additional information on this result.

We now return to the results in Table 6, the bias when ϕ is 0.10 is quite small. Once the sample size is above 25, the estimates are biased by around 2–3% at most and by as little as less than 1%. As we observed in the last set of experiments, increasing ϕ increases the bias. When ϕ is 0.20, the estimates are typically off by 3–7%, still not a large bias. However, once ϕ is 0.50 the bias is substantial. Moreover, across all the values of ϕ,

Table 7 Simulated asymptotic bias in coefficient for lag of Y_t

t	$\hat{\alpha}$
50	0.7285257
75	0.7482767
100	0.7585105
150	0.7697702
200	0.7745333
250	0.7783368
500	0.7844766
1000	0.7882013
5000	0.7905413
10000	0.7906144
20000	0.7906049

Note. 1000 Monte Carlo trials. True α: 0.75.

the bias tends to worsen slightly as ρ increases. Raising ρ from 0.65 to 0.95 adds about 2–3 percentage points to the bias.

Finally, we calculated the bias for the total effect that X_t has on Y_t. We found that when ϕ is 0.00, the total effect tends to be underestimated, but the bias shrinks as t increases. When ϕ is either 0.10 or 0.20, the bias is larger than for the initial effect but typically only by about 2–7%. The bias when ϕ is 0.50, however, is substantial. For all the conditions when ϕ is greater than zero, the estimate of the total effect is overestimated. Please see the online appendix for a table with the full results.

3.4 *Rejection Rates For* $\hat{\beta}$

It would appear that the bias when the residuals in the DGP are modestly autocorrelated is fairly small. But the real question is, how small is small? Generally, the concern is that $\hat{\beta}$ will be too small, making it harder to confirm the effect of X_t on Y_t when it truly exists. In other words, it will be more difficult to reject the null hypothesis that $\hat{\beta} = 0$; as such the probability of making a Type II error will be higher than normal. This implies that we should test how often the null hypothesis for $\hat{\beta}$ is correctly rejected. This rejection rate would then provide us with one benchmark for the magnitude of the bias. That is, we might conclude that the magnitude of the bias is substantial if it frequently leads to incorrect inferences. To do this, we calculated the rejection rate, which is the percentage of times we fail to reject $\hat{\beta} = 0$ out of the 1000 replications.

We find that such false rejections of the null happen rarely, even when the errors of Y_t are highly autocorrelated. Whatever the level of residual autocorrelation in the errors of Y_t, the rejection rate tends to be high only when the sample size is 25. For example, when $\phi = 0.20$ 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 incorrect inferences are 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 t is extremely low, the bias is, in most cases, not large enough to cause an analyst to make incorrect inferences. The weaker the effect of X_t on Y_t, of course, the harder it would be to detect the effect. Please see the online appendix for a table with the full results. Next, we turn to the detection of serial correlation in the estimated residuals.

Table 8 Percentage of positive tests for residual serial correlation

t	ρ	*0.65*	*0.75*	*0.85*	*0.95*
		$\phi = 0.00$			
25		5.6	6.6	6.6	6.0
50		4.8	5.9	4.7	6.1
75		5.8	5.5	4.8	5.9
100		5.6	5.6	4.9	5.4
250		6.1	6.0	6.2	5.4
500		5.8	5.7	5.2	3.7
1000		4.7	5.2	3.9	5.4
		$\phi = 0.10$			
25		6.5	6.2	6.4	5.8
50		8.8	6.5	6.9	6.9
75		11.4	9.1	8.7	9.4
100		13.0	11.9	10.9	11.7
250		25.5	27.1	25.6	28.2
500		46.1	51.2	49.8	53.7
1000		76.7	79.6	81.4	83.5
		$\phi = 0.20$			
25		12.0	9.4	8.6	8.5
50		20.4	20.0	16.0	17.9
75		28.7	29.6	26.8	29.0
100		36.9	35.6	36.1	33.3
250		75.3	75.6	76.4	82.7
500		96.8	97.4	97.3	98.7
1000		100.0	100.0	100.0	100.0
		$\phi = 0.50$			
25		37.2	38.5	37.0	35.6
50		72.7	76.4	74.4	75.9
75		91.7	92.5	92.5	93.7
100		96.9	97.3	97.8	99.0
250		100.0	100.0	100.0	100.0
500		100.0	100.0	100.0	100.0
500		100.0	100.0	100.0	100.0

Note. Results are based on 1000 Monte Carlo replications. Cell entries represent the percentage of time autocorrelation is detected in estimated residuals. Test for autocorrelation is Breusch-Godfrey LM test.

3.5 *Detection of Residual Autocorrelation*

So long as the residuals are not highly autocorrelated, we have shown that estimates of β will exhibit only small amounts of bias. But if the residual autocorrelation is high, an LDV estimated with OLS can produce substantial amounts of bias. We now ask whether we can detect the residual serial correlation that is the source of the bias. If we are able to detect the residual autocorrelation that causes the bias, then we can avoid using an LDV with OLS when we know the estimates are severely biased. As part of our experiments, we tested for residual serial correlation after estimating the LDV model. We now report how often we are able to detect autocorrelation in the estimated residuals. Table 8 contains the results from this test.

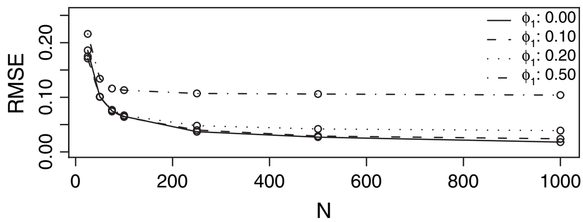

Fig. 3 Comparison of RMSE across values of ϕ.

When the DGP is void of residual autocorrelation, the Breusch-Godfrey LM test for autocorrelation should detect autocorrelated residuals due merely to chance, on average, 5% of the time. We find in Table 8 that this is generally true. There is slight improvement as the sample size grows, but it is modest. This is in direct contrast to the situation in which ϕ is 0.10 or 0.20. Here, the ability to detect the residual autocorrelation is directly contingent on sample size. In both these conditions, it will be fairly difficult to detect the residual autocorrelation with samples of less than 250–500. But it is also under these conditions that we have found that the substantive size of the bias is not problematic. Once ϕ is 0.50 and the sample size is above 25, the residual autocorrelation is usually detectable. In fact, with 100 cases there is only a 2–3% chance of not detecting the residual autocorrelation. This is reassuring. For the condition in which the bias is a serious problem, we can be confident that the residual serial correlation will be detected.

3.6 *Comparing the Results*

The results thus far do not give a precise summary of how the performance of OLS changes as we move from the ideal condition of ϕ being 0.00 to when it rises to 0.10, 0.20, or 0.50. We now compare the overall performance of OLS across the four levels of ϕ. To make the comparison, we hold ρ at 0.85 and then plot the root mean square error (RMSE) across the six different sample sizes for each level of ϕ.[9]

The plot emphasizes how small the difference in overall model performance is when ϕ is either 0.10 or 0.20 as opposed to 0.00. The RMSE is practically indistinguishable for these three values of ϕ. Figure 3 clearly emphasizes that for realistic levels of residual autocorrelation in Y_t, the performance of OLS is not greatly different than under conditions in which OLS is consistent. It is only when ϕ is quite large at 0.50 that the RMSE is noticeably higher.

4 Remarks and Recommendations

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 with OLS. We make several basic points.

First, researchers should be hesitant to use either GLS or OLS with corrected standard errors if one suspects that the process is at all dynamic. Even when the process is weakly dynamic, OLS without a lag was biased, and if the process was strongly dynamic, the bias caused by the specification error in OLS and GLS was dramatic. The probability that a process is at least weakly dynamic is too great to ever use the GLS estimator or OLS without lags given the amount of bias observed in the Monte Carlo analysis.

[9]The RMSE calculation here includes both model parameters: α and β.

Second, if analysts suspect a common factor, they should use an ARMA model. While the performance of the ARMA model was not much better than other alternatives when the DGP only contained a common factor, the ARMA model did much better when the DGP was weakly dynamic. Here, most of the other models fared quite poorly. So it is this robustness that recommends the ARMA model over the alternatives.

Third, if the process was dynamic, OLS with an LDV provided estimates that were superior to the other models or estimators even in the presence of minor residual autocorrelation. In fact, the Monte Carlo evidence underscored the robustness of OLS when used with an LDV on several points. The overall performance of OLS as measured by the RMSE varied little until it reached the highest level of residual serial correlation. Moreover, the autoregressive nature of the X_t variable had little impact, and large samples sizes were not required for good estimates. Even with as few as 50 cases, the OLS estimates were quite good. It was only when ϕ climbed to 0.50 that the bias became substantial.

Our simulations also reveal a surprising and counterintuitive fact about OLS with an LDV. When there is residual serial correlation, the effect of sample size works in the opposite direction that one would expect. That is, OLS produces the best estimates for modest instead of large sample sizes. Given the complications of deriving the small sample properties of OLS with autocorrelated errors and an LDV, analysts have relied on asymptotic derivations of the bias. But we demonstrate that this reliance is misleading since the bias in $\hat{\alpha}$ converges to the asymptotic values as t increases. Thus for the sample sizes most often used in applied work (50–150), one would expect very good estimates for β.

However, the statistical results do offer some caveats. If the model residuals are strongly autocorrelated, including a lag will produce biased estimates. Fortunately, the Breusch-Godfrey LM test almost always detects the residual autocorrelation under these circumstances. As a result, after the estimation of any model with an LDV, the analyst should test that the model residuals are white noise through the use of a LaGrange multiplier test. And Durbin and Watson's d statistic for autocorrelation is not valid when there are lags (of Y_t or X_t) on the right-hand side of the model, since lags in the model bias the d statistic. If the model residuals exhibit significant autocorrelation, the LDV is inappropriate without at least some change in specification, if not a change in estimator. (Greene 2003) outlines a GLS with instrumental variables estimator for lags with common factors.[10]

Finally, analysts must test that the dependent variable is stationary before using OLS with an LDV. Many of the problems that LDVs may cause with budgetary and arms race data probably occur because the data are nonstationary. If the data are nonstationary, OLS with an LDV is clearly incorrect and techniques for cointegrated data should be used instead. Whatever the strengths of LDVs, they are inappropriate with nonstationary data that have not been differenced.

One unanswered question, though, is how does one differentiate between the common factor and dynamic contexts? Here the answers are less certain, as there is no simple test for distinguishing whether the data have a common factor or instead are dynamic. This issue, however, comes down to a theoretical question: Does the past matter for the current values of the process being studied? If the answer is yes, OLS with an LDV is appropriate so long as the stationarity condition holds and the model residuals are not highly autocorrelated. The preponderance of the evidence in both economics and political science is that many if not most cross-temporal processes are dynamic.[11] So if one suspects that history matters, OLS with an LDV model remains a good choice.

[10]This approach may be of little help since it introduces all the difficulties of finding appropriate instruments.

[11]For processes such as budgets and arms races, if they are nonstationary, they are dynamic processes as well, the difference being that there is no decay in the effect of history on the current value of the process.

Appendix

This section of the appendix further explores the patterns of the bias given varying sample sizes. In the tables in the text, the bias tends to increase with the sample size. We found that this is due to the properties of the coefficient of the lag of Y_t. For small samples it will be underestimated, but as the sample size grows, the bias changes direction in a smooth pattern such that the estimate is quite good for intermediate samples but becomes an overestimation that increases for larger samples. For very large samples, the overestimation approaches the asymptotic values. The simulations in Table 7 demonstrate this phenomenon. The simulations for that table were produced with the following DGP:

$$Y_t = \alpha Y_{t-1} + u_t \tag{A.1}$$

where:

$$u_t = \phi u_{t-1} + e_{2t}. \tag{A.2}$$

We set α to 0.75 and ϕ to 0.10. In other results not presented we set ϕ to 0.20 and 0.50, but neither value made any difference. We also reproduced the same result in *Stata*, a different program from the one we used for the simulations reported in this article. We also tested whether this result was caused by the initial value of the time series. Here, for each simulation, we generated $t + 1000$ cases and then dropped the first 1000 values of the time series. Again, we found the same result. The results are also invariant to how we set the seed. We used both fixed seeds and seeds based on system time.

We also investigated the literature for analytical guidance on the result. We found no analytical derivations in the literature for fixed sample sizes when ϕ is above 0.0, as the analytic work calculates the bias asymptotically. This is probably due to the complexity of such a derivation. Even when ϕ is 0.0, the analytic derivations for fixed sample sizes are extremely complex. See (White 1961) and (Phillips 1977) for examples.

References

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

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.

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

Greene, William H. 2003. *Econometric Analysis*, 5th ed. New York: Macmillan.

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. 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.