PA

Lagged Outcomes, Lagged Predictors, and Lagged Errors: A Clarification on Common Factors

Scott J. Cook [1] and Clayton Webb[2]

[1] Associate Professor of Political Science, Department of Political Science, Texas A&M University, College Station, TX 77843, USA. Email: sjcook@tamu.edu, URL: scottjcook.net
[2] Assistant Professor of Political Science, Department of Political Science, University of Kansas, Lawrence, KS 66045, USA. Email: webb767@ku.edu, URL: claytonmwebb.com

Abstract

Debate on the use of lagged dependent variables has a long history in political science. The latest contribution to this discussion is Wilkins (2018, *Political Science Research and Methods*, 6, 393–411), which advocates the use of an ADL(2,1) model when there is serial dependence in the outcome and disturbance. While this specification does offer some insurance against serially correlated disturbances, this is never the best (linear unbiased estimator) approach and should not be pursued as a general strategy. First, this strategy is only appropriate when the data-generating process (DGP) actually implies a more parsimonious model. Second, when this is *not* the DGP—e.g., lags of the predictors have independent effects—this strategy mischaracterizes the dynamic process. We clarify this issue and detail a Wald test that can be used to evaluate the appropriateness of the Wilkins approach. In general, we argue that researchers need to always: (i) ensure models are dynamically complete and (ii) test whether more restrictive models are appropriate.

Keywords: time series, specification testing, lagged dependent variables

Whether and when to use lagged dependent variables (LDVs) has been a long-standing question in political science (Achen 2000; Keele and Kelly 2006). Of particular concern has been the consequence(s) of including a LDV in a model in the presence of residual autocorrelation. Because the LDV has power against error persistence, the coefficient for the LDV will generally be inflated, and the coefficients for (persistent) predictors will be deflated.

In a recent paper, Wilkins (2018) re-engages this question, suggesting that including an additional lag of the outcome and predictor—that is, an ADL(2,1) model[1] —offers leverage against such biases and should be preferred as a more general model specification.[2] While Wilkins (2018) correctly notes that a time-lagged error can be re-expressed as a time lag of the outcome and predictor (see Sargan 1964), we are concerned that other aspects of this discussion invite confusion.

First, such a model only suffices insofar, as it is dynamically complete. That is, one first needs to ensure that any model fully characterizes the dependence in the series, imposing only those restrictions supported by the data (Hendry 1995). Second, even when the ADL(2,1) model is sufficient, the strategy offered by Wilkins (2018) *assumes* that the underlying data-generating process (DGP) is a first-order partial adjustment (i.e., PA[1]) process (familiarly known as the LDV model) with autocorrelation in the residuals. If the DGP is actually a more general ADL(2,1) process—with meaningful effects of x_{t-1} and y_{t-2}—Wilkins's approach mischaracterizes the dynamic process, inviting incorrect interpretations of the model coefficients and the long-run multiplier (LRM).

---

1 Here we use standard notation for the autoregressive distributed lag (ADL) model of order p and q, ADL(p,q), where p indicates lags of y and q lags of x.

2 This paper has already proved influential, as it has been cited more than 80 times to date.

Rather than proceed by assumption, we argue analysts should test whether these parameter restrictions are supported by their data.[3]

To this end, we identify the nonlinear common factor restriction required to support Wilkins's interpretation, suggest an associated Wald test to compare his proposed specification to alternative models, and demonstrate its efficacy via stochastic simulation. We caution researchers against privileging any single specification by default. Instead, we advocate that they undertake a general-to-specific specification search using a higher-order ADL(p,q) model, test whether lags in the structural equations are proxying for residual autocorrelation, and properly calculate quantities of interest.

1 Model Equivalence and Common Factor Restrictions

As first illustrated by Sargan (1964), time-lagged realizations of a model's structural terms—the outcome and its predictors—can be used to proxy for time-lagged realizations of its stochastic error process. Consider

$$y_t = x_t \beta + u_t, \text{ where } u_t = \rho u_{t-1} + e_t,$$

y and x are covariates measured at time t, and u_t is an autoregressive error term, which is a function of prior realizations (via ρ) and contemporaneous white noise residuals e_t. Using the familiar backshift operator L and rearranging terms, this process can be expressed as an ADL(1,1):

$$y_t = x_t \beta + u_t, \text{ where } u_t = \rho u_{t-1} + e_t, \tag{1a}$$

$$y_t = x_t \beta + (1 - \rho L)^{-1} e_t, \tag{1b}$$

$$(1 - \rho L) y_t = (1 - \rho L) x_t \beta + e_t, \tag{1c}$$

$$y_t = \rho y_{t-1} + x_t \beta - \rho x_{t-1} \beta + e_t, \tag{1d}$$

$$y_t = \alpha y_{t-1} + x_t \beta_1 + x_{t-1} \beta_2 + e_t. \tag{1e}$$

This demonstrates how an ADL(1,1) model captures the error persistence from a PA(1) process using lags of observed variables.[4] This approach is widely known and has been discussed at length in the time-series literature (Hendry and Mizon 1978; Sargan 1964, 1980). Historically, this specification was valuable, because ADL models could be estimated using ordinary least squares.

Yet, Sargan (1980) and others cautioned that this approach has limitations. First, the equivalence is only obtained if the implied common factor restrictions of the reduced-form parameters are valid. That is, Equation (1e) can only be interpreted as a static model with autocorrelation in the residuals if $\beta_2 = -\alpha \beta_1$, allowing the simplification undertaken in the step from Equation (1d) to (1e).[5] If x_{t-1} has independent effects, these restrictions are not met, and the estimator is biased (Sargan 1964).[6] Second, even when this restriction is satisfied, the ADL(1,1) model is inefficient, since the static model has fewer parameters to be estimated (Hendry and Mizon 1978).

---

3 Arbitrating between the PA(1) and ADL(2,1) models presumes that the analyst has first classified their data as stationary (Webb, Linn, and Lebo 2020), specified a balanced model where all the dynamic features of the regressand (order of integration, trend, seasonality, etc.) are accounted for (Enns and Wlezien 2017; Granger 1990; Pickup and Kellstedt 2020), and determined the ADL(2,1) model is parsimonious and dynamically complete. If instead the data are nonstationary, conventional hypothesis testing procedures will not be appropriate for the ADL or PA (Enns *et al.* 2016), instead researchers should use the critical value bounds developed by Webb, Linn, and Lebo (2019).

4 As we demonstrate in the Online Appendix, this equivalence can also be demonstrated without the use of the backshift operator.

5 While it is well known, we derive this common factor restriction in the Online Appendix.

6 To clarify, we use "independent effects" to indicate that the mean of the outcome is a function of these inputs and that they are not simply a proxy for stochastic error persistence. Put differently, in their absence, the model would suffer from omitted variables bias, and not simply inefficiency.

In a recent piece, Wilkins (2018) uses similar reasoning to argue that an ADL(2,1) model can be used to estimate a PA(1) process with autocorrelation in the residuals, thereby resolving the issue of LDVs raised in Achen (2000). As above, this can be achieved as follows:

$$y_t = \alpha y_{t-1} + x_t \beta + u_t, \text{ where } u_t = \rho u_{t-1} + e_t, \tag{2a}$$

$$y_t = \alpha y_{t-1} + x_t \beta + (1 - \rho L)^{-1} e_t, \tag{2b}$$

$$(1 - \rho L) y_t = (1 - \rho L) \alpha y_{t-1} + (1 - \rho L) x_t \beta + e_t, \tag{2c}$$

$$y_t = (\rho + \alpha) y_{t-1} - \rho \alpha y_{t-2} + x_t \beta - \rho x_{t-1} \beta + e_t, \tag{2d}$$

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + x_t \beta_1 + x_{t-1} \beta_2 + e_t. \tag{2e}$$

From this, Wilkins (2018) argues in favor of the ADL(2,1) model as a more general approach, since, unlike the LDV model considered by Achen (2000), the ADL(2,1) model is robust to this additional error persistence.

While Wilkins's strategy echoes that of Sargan and others, he is silent on their cautions. Most importantly, he does not discuss the common factor restriction required for this model equivalence to hold. As we detail in the Online Appendix, an ADL(2,1) process can reduce to a PA(1) process with autocorrelation in the residuals if, and only if, $\beta_2^2 + \beta_1 \beta_2 \alpha_1 - \alpha_2 \beta_1^2 = 0$. When this does not hold, it implies that the second-order lag of the outcome or the first-order lag of the predictor has true, independent effects on the contemporaneous outcome. That is, they have an effect above and beyond proxying for the lagged stochastic error, so interpreting these estimates as such can lead to a mischaracterization of the dynamic process. For example, in a regression of presidential approval (y) on consumer sentiment (x), the Wilkins (2018) approach assumes that there is no lagged effect of consumer sentiment (x_{t-1}) and no second-order autocorrelation in approval (y_{t-2}), with the coefficients of these covariates understood to exclusively reflect error persistence. When these assumptions are invalid—e.g., lagged consumer sentiment impacts contemporaneous approval— this interpretation of the parameters is not supported.

Not only would this mischaracterize specific coefficient estimates, but any marginal effects obtained from these parameters will also be incorrect. For example, the LRM for the effect of x on y for an ADL(2,1) model is

$$\frac{\beta_1 + \beta_2}{1 - \alpha_1 - \alpha_2}, \tag{3}$$

which Wilkins (2018) argues can recover the LRM for the PA(1) process as

$$\frac{\beta - \beta \rho}{1 - \alpha - \rho + \alpha \rho} = \frac{\beta(1 - \rho)}{(1 - \alpha)(1 - \rho)} = \frac{\beta}{1 - \alpha}, \tag{4}$$

where the reduced-form coefficients from the ADL(2,1) model are substituted in for their functional relations in the PA(1) process with autocorrelation in the residuals.[7] However, Equations (3) and (4) will *only* be equal if the common factor restriction—i.e., $\beta_2^2 + \beta_1 \beta_2 \alpha_1 - \alpha_2 \beta_1^2 = 0$—holds.[8] This restriction can be satisfied, given the right set of reduced-form coefficients; however, it should not be assumed. Only in stylized cases (e.g., $\alpha_2 = 0$ and $\beta_2 = 0$) will it be easy for researchers to easily determine whether the restriction is satisfied, and more often, it will entail complicated combina-

---

7 Wilkins (2018) also argues that the LRMs should be calculated with dynamics of x in the denominator. Since there is some controversy on this issue and it is not the focus of our paper, we simply assume that x is a static series.

8 An alternative formulation of the LRM under Wilkins's (2018) assumptions is $\frac{\hat{\beta}_1}{1 - (\hat{\beta}_2/\hat{\beta}_1 + \hat{\alpha}_1)}$, where the ADL(2,1) coefficients are used to produce the PA(1) LRM, $\frac{\beta}{1-\alpha}$. We make use of this in our simulations in Section 2.1.

tions of the coefficients (e.g., $\beta_1 = 5, \beta_2 = 2, \alpha_1 = 0.5$, and $\alpha_2 = 0.36$).[9] When this restriction is not satisfied, Equation (4) mischaracterizes the LRM, inaccurately reflecting both the direct effect (i.e., $\beta_1 + \beta_2 \neq \beta$) and the persistence (i.e., $1 - \alpha_1 - \alpha_2 \neq 1 - \alpha$).[10] Using our earlier example of presidential approval, misinterpreting the effect of lagged consumer sentiment as error persistence impacts not only (β_2), but propagates through the LRM to bias our understanding of the effect of consumer sentiment on approval more generally.

As such, it is important for researchers to determine whether they have a traditional ADL(2,1) process, i.e., Scenario A, where the structural lags have independent effects, or a Wilkins (2018) ADL(2,1) process, i.e., Scenario B, where these lags are proxies for the stochastic error process. Even when Wilkins's interpretation is correct (Scenario B), and an ADL(2,1) model can rightly be reparameterized as a PA(1) model, Wilkins's estimation strategy is inefficient, because it estimates one more reduced-form parameter than necessary to identify the structural equations. Moreover, one loses an additional year of data, since y_{t-2} is used as an input. Neither of these matters asymptotically, but efficiency losses are greater in shorter series. This is especially important in time-series analysis, where sample coefficient estimates are used as inputs for additional quantities of interest (e.g., the LRM and impulse response functions). For these nonlinear combinations of coefficients, slight efficiency losses may have severe consequences.

In the next section, we use simulated data to quantify the costs of using the Wilkins (2018) strategy when its assumptions are not maintained. Given these costs, we also demonstrate how researchers can use a simple Wald test—evaluating $\beta_2^2 + \beta_1 \beta_2 \alpha_1 - \alpha_2 \beta_1^2 = 0$—to distinguish between Scenarios A and B in Section 2.2.

2 Simulations

We use simulations to evaluate the bias in the LRM under incorrect assumptions and the efficacy of our proposed Wald test. The outcome, y, is generated:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_1 x_t + \beta_2 x_{t-1} + u_t, \tag{5}$$

where $x_t \sim N(0, 1)$ and $u_t = \rho u_{t-1} + e_t$ with $e_t \sim N(0, 1)$. We hold the contemporaneous effect fixed, $\beta_1 = 5$, varying the strength of the lagged predictor via $\beta_2 = \{0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50\}$, the lagged outcomes via $\alpha_1 = \{0.00, 0.20, 0.40\}$ and $\alpha_2 = \{0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50\}$, and the residual autocorrelation via $\rho = \{0.00, 0.20, 0.40\}$. For each combination of parameters, we generate 1,000 simulated data sets with sample sizes of $T = 50, 100, 200, 1000$.[11]

2.1 LRM Bias Associated with Incorrect Specification

One traditionally uses Equation (3) to calculate the LRM from the estimates of an ADL(2,1) model. For comparison, we directly calculate the LRM interpretation proposed by Wilkins (2018) given in Equation (4), which uses the ADL(2,1) estimates to capture a PA(1) process with autocorrelation. We calculate the bias for the LRM as the difference between the true LRM, $\frac{\beta_1 + \beta_2}{1 - \alpha_1 - \alpha_2}$, and the LRM suggested by the Wilkins (2018) strategy, $\frac{\hat{\beta}_1}{1 - (\hat{\beta}_2/\hat{\beta}_1 + \hat{\alpha}_1)}$, where the coefficients from the ADL(2,1) model are substituted to represent the LRM for the PA(1) process, $\frac{\beta}{1 - \alpha}$. If the data support the

---

9 Fully evaluating the latter case would also require uncertainty estimates beyond those typically reported with individual coefficient estimates, as one would need to estimate the variance of $\beta_2^2 + \beta_1 \beta_2 \alpha_1 - \alpha_2 \beta_1^2$.

10 One could argue that since (3) produces unbiased LRMs for general ADL(2,1) processes, this should just be preferred. However, as we note throughout, under the conditions given by Wilkins (2018), using the ADL(2,1) actually produces efficiency losses and risks inferential errors on the coefficients on x_{t-1} and y_{t-2}. This is why researchers should use tests like the one given here to arrive at the correct model specification first and then calculate the LRM directly from the coefficients of that model.

11 The first three sample sizes are common in applied work, while the last approximates asymptotics.
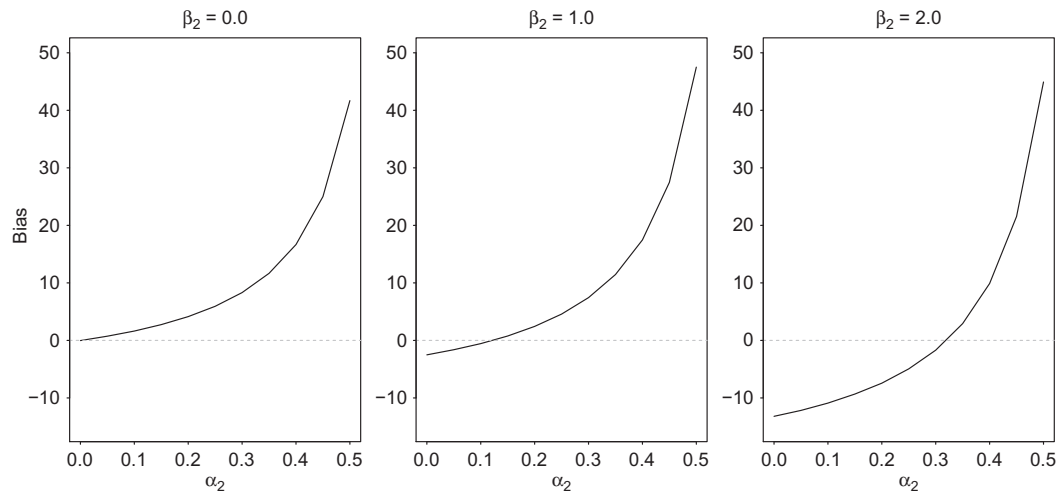
Figure 1. LRM bias over values of α_2. Notes: Median bias is computed based on the difference between the true LRM (Equation (3)) and the LRM restrictions suggested by Wilkins (2018). Results are shown for $T = 50$, $\alpha_1 = 0.4$, and $\rho = 0.4$.

Wilkins (2018) interpretation, these two will be equal to one another, as in Equation (4). If the data do not support the Wilkins (2018) interpretation, the difference between the two will reflect the extent of the bias from reinterpreting the ADL(2,1) estimates as if they had been produced by a PA(1) process with residual autocorrelation.

Because these LRMs are nonlinear in parameters, the resultant biases will also change in a nonlinear fashion. In Figure 1 we focus on the consequences of changes to α_2 (x-axis), holding β_2 at fixed values (0.0 in panel 1, 1.0 in panel 2, and 2.0 in panel 3). The curves show the median bias in the LRMs (y-axis). In each panel, there is only one set of conditions where the bias is equal to zero. In the first panel, there is no bias when $\beta_2 = 0$ and $\alpha_2 = 0$. When $\beta_2 \neq 0$ in the second and third panels, the LRMs are biased except for the conditions, where the value of α_2 exactly offsets β_2, ($\alpha_2 = 0.12$ and $\beta_2 = 1$) and ($\alpha_2 = 0.32$ when $\beta_2 = 2$), respectively. The bias increases as α_2 increases beyond these thresholds. In sum, where there is a *true* effect of y_{t-2}, the LRM proposed by Wilkins (2018) is biased. However, in applied data settings, we would not know whether this bias attenuates or inflates the LRM, because it is a nonlinear combination of several parameters.

We show similar results in Figure 2, where we focus on changes to β_2 (x-axis) while holding α_2 at fixed values (0.0 in panel 1, 0.3 in panel 2, and 0.5 in panel 3). As before, in the first plot, the LRM bias is 0 when $\beta_2 = 0$ and $\alpha_2 = 0$. Equations (3) and (4) are equivalent in this condition. As the value of β_2 increases along the x-axis, the LRM implied by the Wilkins (2018) strategy underestimates the true value of the LRM at an increasing rate. The same pattern exists in the second ($\alpha_2 = 0.3$) and third ($\alpha_2 = 0.5$) plots, but the y-intercept for the bias ($\beta_2 = 0$) is different in both cases.

The results presented in Figures 1 and 2 demonstrate the potential problems with assuming the restrictions proposed by Wilkins (2018). If the true DGP is an ADL(2,1), the LRM formula proposed by Wilkins (2018) is a biased estimator. Moreover, in applied research, the direction and magnitude of the bias are difficult to predict, because the nature of the bias depends on the values of α_1, α_2, β_1, and β_2. As such, researchers cannot confidently assume the effects are being under- or overestimated.

2.2 A Wald Test for the ADL(2,1) Against a PA(1) with Autocorrelation

The results presented in the last section highlight the perils associated with incorrectly calculating the LRM for an ADL(2,1) process as though it were generated by a PA(1) process. On the other hand, Wilkins (2018) demonstrates the biases risked by failing to impose these restrictions when the true
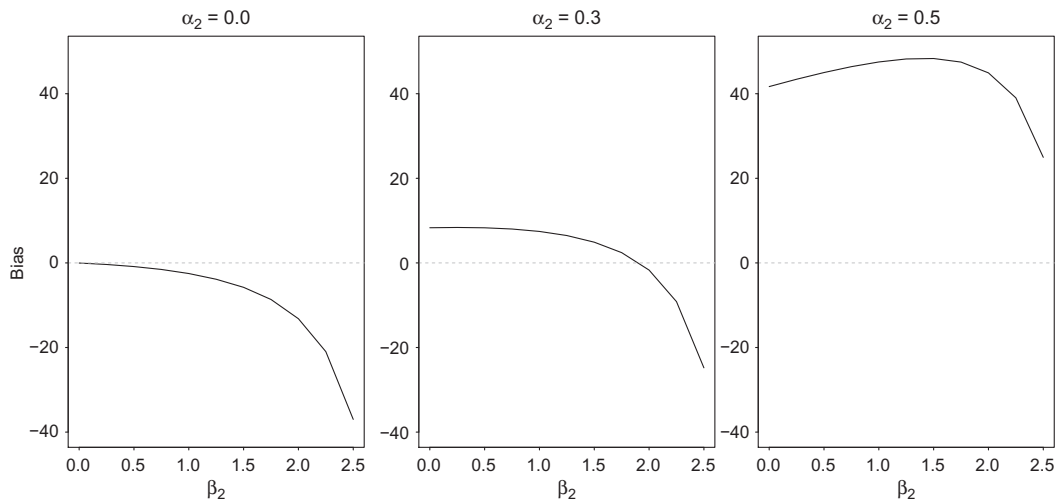
Figure 2. LRM bias over values of β_2. Notes: Median bias is computed based on the difference between the true LRM (Equation (3)) and the LRM restrictions suggested by Wilkins (2018). Results are shown for $T = 50$, $\alpha_1 = 0.4$, and $\rho = 0.4$.

DGP is a PA(1) process with autocorrelation in the residuals. In either case, proceeding purely from assumption is a risky strategy.

In Section 1, we discussed a possible test to evaluate the restrictions assumed by Wilkins (2018). This draws from a strategy outlined by Sargan (1964, 1980), which demonstrates how Wald tests can be used to compare a wide range of time-series specifications. Specifically, we test whether estimated ADL(2,1) coefficients are consistent with a PA(1) process with residual autocorrelation by testing

$$\beta_2^2 + \beta_1 \beta_2 \alpha_1 - \alpha_2 \beta_1^2 = 0. \tag{6}$$

This nonlinear Wald test is χ^2 distributed with 1 (the number of restrictions being tested) degree of freedom. The null hypothesis is that the ADL(2,1) is indistinguishable from a PA(1) with residual autocorrelation. The alternative hypothesis is that the data were generated by an alternative ADL(2,1) process, where y_{t-2} and x_{t-1} have independent effects. As such, this test enables researchers to evaluate whether their data are consistent with the interpretation suggested by Wilkins (2018) or not, avoiding the biases demonstrated above.

We demonstrate the efficacy of the proposed test using the simulations described in the previous section.[12] The results are presented in Table 1, which has four panels, one for each of the sample sizes. Each element in each panel gives the rejection rate for the respective combination of α_1, α_2, and β_2.[13]

Looking first at the rejection rates when $\alpha_2 = \beta_2 = 0$ (in *italics*), we demonstrate the size of the test. These rejection rates are, approximately, the expected 0.05, with somewhat worse performance in small samples. Demonstrating the power of the test is not straightforward, since, as noted above, increases to the individual parameters do not always increase the total of $\beta_2^2 + \beta_1 \beta_2 \alpha_1 - \alpha_2 \beta_1^2$. Therefore, we focus on a particular case ($\alpha_1 = 0.4$ and $\alpha_2 = 0.0$), where $\beta_2^2 + \beta_1 \beta_2 \alpha_1 - \alpha_2 \beta_1^2$ strictly increases (0.00, 0.56, 1.25, 2.06, and 3.00) as β_2 increases (0.00, 0.25, 0.50, 0.75, and 1.00), thereby giving us clearer insight into the power of the test.

---

12 As before, we present results for $\rho = 0.4$ in the main text, with results for $\rho = 0.0$ and $\rho = 0.2$ in the Online Appendix.
13 Recall that the tested condition, i.e., Equation (6), is not linearly (or even monotonically) increasing in these individual parameters. Therefore, to aid interpretation, we also calculate and report the value of the sample test statistic implied by various combinations of parameters. These results are given in the Online Appendix.

Table 1. Wald test for ADL(2,1) against PA(1), $\rho = 0.4$.

	$\beta_2 =$	0.00	0.25	0.50	0.75	1.00	0.00	0.25	0.50	0.75	1.00
				T = 50					*T = 100*		
$\alpha_1 = 0.0$	$\alpha_2 = 0.0$	*0.07*	0.07	0.08	0.12	0.24	*0.06*	0.06	0.08	0.18	0.41
	$\alpha_2 = 0.2$	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	$\alpha_2 = 0.4$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\alpha_1 = 0.2$	$\alpha_2 = 0.0$	*0.07*	0.09	0.18	0.37	0.66	*0.05*	0.10	0.29	0.64	0.94
	$\alpha_2 = 0.2$	1.00	1.00	1.00	0.98	0.91	1.00	1.00	1.00	1.00	1.00
	$\alpha_2 = 0.4$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\alpha_1 = 0.4$	$\alpha_2 = 0.0$	*0.06*	0.14	0.35	0.69	0.90	*0.06*	0.20	0.60	0.95	1.00
	$\alpha_2 = 0.2$	1.00	1.00	0.98	0.89	0.58	1.00	1.00	1.00	1.00	0.91
	$\alpha_2 = 0.4$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
				T = 200					*T = 1000*		
$\alpha_1 = 0.0$	$\alpha_2 = 0.0$	*0.06*	0.06	0.09	0.28	0.68	*0.05*	0.05	0.23	0.86	1.00
	$\alpha_2 = 0.2$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_2 = 0.4$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\alpha_1 = 0.2$	$\alpha_2 = 0.0$	*0.06*	0.12	0.47	0.90	1.00	*0.05*	0.41	0.99	1.00	1.00
	$\alpha_2 = 0.2$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_2 = 0.4$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\alpha_1 = 0.4$	$\alpha_2 = 0.0$	*0.05*	0.32	0.88	1.00	1.00	*0.05*	0.89	1.00	1.00	1.00
	$\alpha_2 = 0.2$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_2 = 0.4$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Notes: Rejection rates are the proportion of the 1,000 simulations where $\beta_2^2 + \beta_1\beta_2\alpha_1 - \alpha_2\beta_1^2 = 0$. The Wald tests are χ^2 distributed with $q = 1$ degrees of freedom.

First, for each sample size, the power of the test is strictly increasing in the magnitude of the population parameter—that is, as we move to the right across the table in this row. In the asymptotic sample size ($T = 1,000$), for example, we see that corresponding rejection rates to each condition are: 0.05, 0.89, 1.00, 1.00, 1.00. Encouragingly, we see that the size of the test is exact, and the power of the test quickly increases to unity. Comparing these conditions across T demonstrates the importance of sample size. In the $T = 50$ case, for example, the rejection rates drop to: 0.06, 0.14, 0.35, 0.69, 0.90. This indicates, as one would expect, that the magnitudes of the parameters will need to be larger to discriminate between models when sample sizes are small.

The results presented in this section demonstrate that the test proposed by Sargan (1964) to distinguish static processes with residual autocorrelation from ADL(1,1) processes can be extended to PA(1) processes with residual autocorrelation and ADL(2,1) processes. While an ADL(2,1) model *can* be used to approximate a PA(1) process with serially correlated errors, one cannot assume that all ADL(2,1) models are simply capturing dynamics in the error process. Even small coefficients can produce large differences in the two models. As such, we offered a test for analysts to distinguish the two processes.

3 Discussion

Dynamic specification is critical to sound inference. However, accurately specifying model dynamics is complicated, because (a) theory is usually silent on the specific structure of long-run relationships and (b) we typically rely on data that are not collected with our specific hypotheses in mind. Given these challenges, researchers have long-sought, single, plug-and-play models that

can be used to ensure results are not a consequence of mismodeled dynamics. These efforts are misguided, however, as there is no single best model that can be applied in all conditions. Minimally, researchers need to first consider whether their data are stationary (Webb, Linn, and Lebo 2020), and then determine whether their estimated models are balanced (Granger 1990) and dynamically complete (Hendry 1995). Only then should the specific model specification concerns discussed here be taken up.

We demonstrate that the Wilkins's (2018) ADL(2,1)-as-PA(1) with autocorrelation approach is only appropriate under a restrictive set of assumptions about the reduced-form parameters of the ADL(2,1). When these conditions are not met, this approach risks misunderstanding the dynamic process and produces biased quantities of interest, such as the LRM. To avoid this, we detail a test that can be used to determine whether the conditions assumed by Wilkins (2018) are satisfied. We note, however, that the conditions highlighted by Wilkins (2018) suggest a more parsimonious model is appropriate. In general, we argue that testing whether lagged systematic terms, as in ADL(p,q) models, are proxying for error persistence is sound practice, as it helps to avoid over-parameterized models and possible misattribution of coefficient effects. The results presented in Table 1 highlight that the test generally performs well, helping researchers to determine whether their ADL(2,1) coefficients are indistinguishable from a PA(1) process with residual autocorrelation, or indicate a more general ADL(2,1) process.

While our discussion and simulations above are limited to a specific case where the analyst is arbitrating between two well-defined DGPs, applied researchers are likely to face less clear-cut choices. Time-series analyses are bedeviled by a number of practical challenges including more complex dynamic processes, inappropriate sampling and aggregation, and under-powered tests. Despite this, the strategy we articulate here can, and should, be incorporated as part of standard practice. First, analysts should begin with a plausible general model that reflects what their theory and pretesting tell them about their data and test restrictions on this model to arrive at a dynamic specification that is simultaneously parsimonious and dynamically complete (Hendry 1995). Second, since lagged systematic terms have power against error persistence, researchers should use the test discussed above in conjunction with traditional testing-down approaches. Finally, analysts should draw complete inferences from their models by calculating the LRM and other quantities of interest (De Boef and Keele 2008). We also caution researchers against overinterpreting coefficients for direct effects of lagged covariates. As demonstrated both here and in Wilkins (2018), these terms have power against stochastic processes, which invites misinterpretation, as the coefficients may reflect systematic effects, stochastic effects, or both.

Acknowledgement

Data Availability Statement

The replication materials for this paper can be found at Webb and Cook (2020).

Supplementary Material

For supplementary material accompanying this paper, please visit https://dx.doi.org/10.1017/pan.2020.53.

References

Achen, C. H. 2000. "Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables." Working Paper. Department of Political Science and Institute for Social Research, University of Michigan, Ann Arbor, Michigan. http://www-personal.umich.edu/~franzese/Achen.2000.LDVstealingExplanPower.pdf.

De Boef, S., and L. Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52(1):184–200.

Enns, P. K., N. J. Kelly, T. Masaki, and P. C. Wohlfarth. 2016. "Don't Jettison the General Error Correction Model Just Yet: A Practical Guide to Avoiding Spurious Regression with the GECM." *Research & Politics* 3(2):2053168016643345.

Enns, P. K., and C. Wlezien. 2017. "Understanding Equation Balance in Time Series Regression." *The Political Methodologist* 24(2):2–12.

Granger, C. 1990. "Where Are the Controversies in Econometric Methodology." In *Modeling Economic Series*, 1–28. New York: Oxford University Press.

Hendry, D. 1995. *Dynamic Econometrics*. New York: Oxford University Press.

Hendry, D. F., and G. E. Mizon 1978. "Serial Correlation as a Convenient Simplification, Not a Nuisance." *The Economic Journal* 88(351):549–563.

Keele, L., and N. J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14(2):186–205.

Pickup, M., and P. Kellstedt. 2020. "Equation Balance in Time Series Analysis: What It Is and How to Apply It." Working Paper (January 28, 2020). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3526534.

Sargan, J. D. 1964. "Wages and Prices in the United Kingdom: A Study in Econometric Methodology." *Econometric Analysis for National Economic Planning* 16:25–54.

Sargan, J. D. 1980. "Some Tests of Dynamic Specification for a Single Equation." *Econometrica* 48(4):879.

Webb, C., and S. J. Cook. 2020. "Replication Data for: Lagged Outcomes, Lagged Predictors, and Lagged Errors: A Clarification on Common Factors." https://doi.org/10.7910/DVN/GWSZV3, Harvard Dataverse, V1.

Webb, C., S. Linn, and M. Lebo. 2019. "A Bounds Approach to Inference Using the Long Run Multiplier." *Political Analysis* 27(3):281–301.

Webb, C., S. Linn, and M. Lebo. 2020. "Beyond the Unit Root Question: Uncertainty and Inference." *American Journal of Political Science* 64(2):275–292.

Wilkins, A. S., 2018. "To Lag or Not to Lag?: Re-Evaluating the Use of Lagged Dependent Variables in Regression Analysis." *Political Science Research and Methods* 6(2):393–411.