Online Appendix:
Lagged outcomes, lagged predictors, and lagged errors

Scott J. Cook* Clayton Webb†

*Assistant Professor of Political Science, Department of Political Science, Texas A&M University, College Station, TX 77843. Email: sjcook@tamu.edu, URL: scottjcook.net

†Assistant Professor of Political Science, Department of Political Science, University of Kansas, Lawrence, KS 66045. Email: webb767@ku.edu

A Deriving the Common Factor Restrictions

The equivalence between a static process with residual autocorrelation and an ADL(1,1) process is well known (Sargan, 1964), for completeness we re-derive that here where L is the familiar back-shift operator (i.e., $Ly_t = y_{t-1}$):

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

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

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

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

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

That is, an ADL(1,1) process is equivalent to a static process with residual autocorrelation if the reduced-form parameters in (1e) satisfy $\beta_2 + \alpha\beta_1 = 0$, often called a common factor restriction. To see this, simply re-express the reduced-form parameters in (1e) as the structural parameters in (1d):

$$\alpha = \rho \tag{2a}$$

$$\beta_1 = \beta \tag{2b}$$

$$\beta_2 = -\rho\beta \tag{2c}$$

Given (2a) and (2b) we can re-write (2c) entirely as a function of the reduced-form parameters, $\beta_2 = -\alpha\beta_1$, thereby permitting a non-linear Wald test – wherein the null hypothesis is that the process is a static model with autocorrelated residuals.

Extending this to the case considered by Wilkins (2018), we re-rexpress a PA(1) process with autocorrelated residuals as an ADL(2,1) process:

$$y_t = \alpha y_{t-1} + x_t\beta + u_t, \text{ where } u_t = \rho u_{t-1} + e_t \tag{3a}$$

$$y_t = \alpha y_{t-1} + x_t\beta + (1-\rho L)^{-1} e_t \tag{3b}$$

$$(1-\rho L) y_t = (1-\rho L)\alpha y_{t-1} + (1-\rho L) x_t\beta + e_t \tag{3c}$$

$$y_t = (\rho + \alpha) y_{t-1} - \rho\alpha y_{t-2} + x_t\beta - \rho x_{t-1}\beta + e_t \tag{3d}$$

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + x_t\beta_1 + x_{t-1}\beta_2 + e_t \tag{3e}$$

Here a similar, if more complicated, common factor restriction of the model equivalence can also be obtained. As before, re-expressing the reduced-form parameters in (3e) as the structural parameters in (3d) gives:

$$\beta_1 = \beta \tag{4a}$$

$$\beta_2 = -\rho\beta \tag{4b}$$

$$\alpha_1 = \rho + \alpha \tag{4c}$$

$$\alpha_2 = -\rho\alpha \tag{4d}$$

where using the right-hand side of (4a) and re-arranging terms in (4b) we can solve for $\rho = -\frac{\beta_2}{\beta_1}$. Given ρ we can now solve (4c) and re-arrange to give $\alpha = \alpha_1 + \frac{\beta_2}{\beta_1}$. Finally, we can re-express (4d) entirely as a function of reduced-form parameters as $\beta_2^2 + \beta_1\beta_2\alpha_1 - \alpha_2\beta_1^2 = 0$, permitting a Wald test of the model equivalence. As standard for non-linear Wald tests, the variance is calculated via the delta method and the test statistic is distributed χ^2.

As noted in footnote 4 in the main text, the use of the back-shift operator L is not strictly necessary to demonstrate the equivalence between these models. Here we reproduce Equations 1 and 2 of the main text without it. First, as in Equation 1, we can re-express a static process with residual autocorrelation as an ADL(1,1) process:

$$y_t = x_t\beta + u_t, \text{ where } u_t = \rho u_{t-1} + e_t, \tag{5a}$$
$$y_t = x_t\beta + \rho u_{t-1} + e_t, \text{ where } u_{t-1} = y_{t-1} - x_{t-1}\beta \tag{5b}$$
$$y_t = x_t\beta + \rho(y_{t-1} - x_{t-1}\beta) + e_t \tag{5c}$$
$$y_t = \rho y_{t-1} + x_t\beta - \rho x_{t-1}\beta + e_t \tag{5d}$$
$$y_t = \alpha y_{t-1} + x_t\beta_1 + x_{t-1}\beta_2 + e_t. \tag{5e}$$

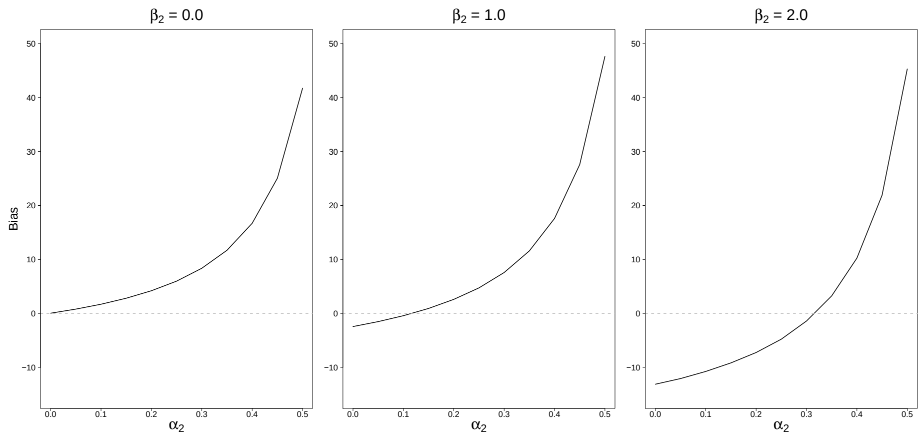
Second, as in Equation 2, we can re-express a PA(1) process with residual autocorrelation as an ADL(2,1) process:

$$y_t = \alpha y_{t-1} + x_t\beta + u_t, \text{ where } u_t = \rho u_{t-1} + e_t \tag{6a}$$
$$y_t = \alpha y_{t-1} + x_t\beta + \rho u_{t-1} + e_t, \text{ where } u_{t-1} = y_{t-1} - \alpha y_{t-2} - x_{t-1}\beta \tag{6b}$$
$$y_t = \alpha y_{t-1} + x_t\beta + \rho(y_{t-1} - \alpha y_{t-2} - x_{t-1}\beta) + e_t \tag{6c}$$
$$y_t = (\rho + \alpha)y_{t-1} - \rho\alpha y_{t-2} + x_t\beta - \rho x_{t-1}\beta + e_t \tag{6d}$$
$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + x_t\beta_1 + x_{t-1}\beta_2 + e_t. \tag{6e}$$

B Simulation Results

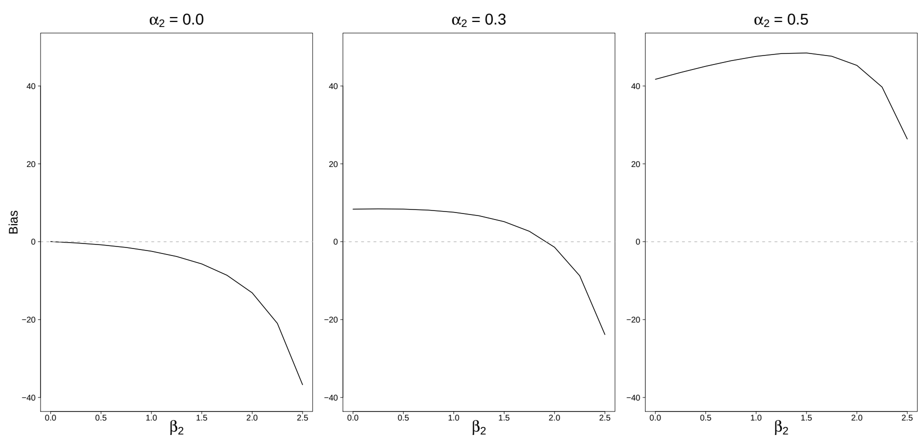
Figures 1 and 2 of the manuscript show the LRM bias when the autoregressive parameter in the residual process ρ is set to 0.4. Figures 1 and 2 in the Appendix, below, shows the corresponding LRM bias when $\rho = 0.0$ and Figures 3 and 4 shows the bias when $\rho = 0.2$.

Figure 1: LRM Bias over values of α_2, $\rho = 0.0$



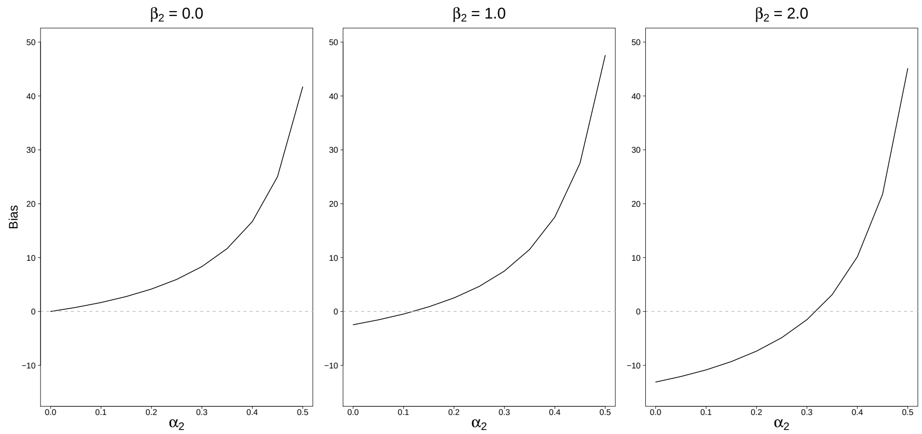
Note: Median bias is computed based on the difference between the true LRM and the LRM restrictions proposed by Wilkins (2018). Results are shown for T = 50, $\alpha_1 = .4$, and $\rho = 0$.

Figure 2: LRM Bias over values of β_2, $\rho = 0.0$



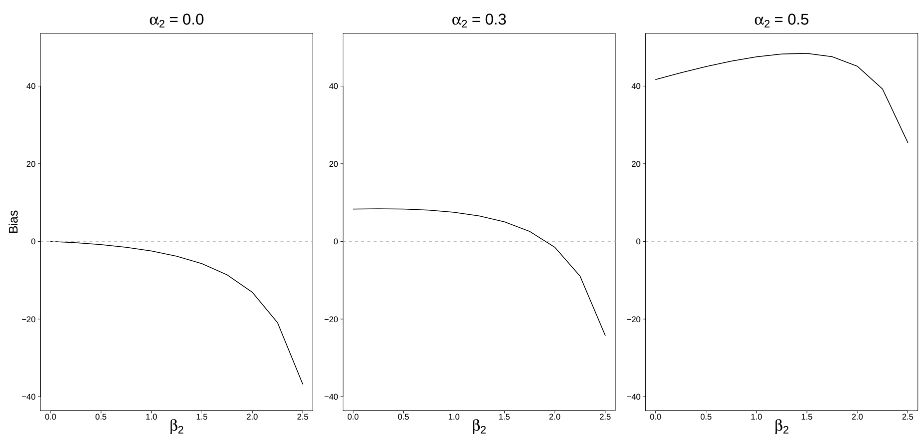
Note: Median bias is computed based on the difference between the true LRM and the LRM restrictions proposed by Wilkins (2018). Results are shown for T = 50, $\alpha_1 = .4$, and $\rho = 0$.

Figure 3: LRM Bias over values of α_2, $\rho = 0.2$



Note: Median bias is computed based on the difference between the true LRM and the LRM restrictions proposed by Wilkins (2018). Results are shown for T = 50, $\alpha_1 = .4$, and $\rho = 0.2$.

Figure 4: LRM Bias over values of β_2, $\rho = 0.2$



Note: Median bias is computed based on the difference between the true LRM and the LRM restrictions proposed by Wilkins (2018). Results are shown for T = 50, $\alpha_1 = .4$, and $\rho = 0.2$.

In Table 1 in the main text, we present the power of our Wald test under a variety of parameter combinations for α_1, α_2, and β_2. Here we report the true population parameter $\beta_2^2 + \beta_1\beta_2\alpha_1 - \alpha_2\beta_1^2$ for each combination of these experimental conditions. Our results in the main paper demonstrate how often our test rejects the null hypothesis using samples generated from these conditions.

Table 1: Sample Test Statistic for ADL(2,1) against PA(1) with residual autocorrelation

	β_2 =	0.00	0.25	0.50	0.75	1.00
α_1 = 0.0						
	α_2 = 0.0	0.00	0.06	0.25	0.56	1.00
	α_2 = 0.2	-5.00	-4.94	-4.75	-4.44	-4.00
	α_2 = 0.4	-10.00	-9.94	-9.75	-9.44	-9.00
α_1 = 0.2						
	α_2 = 0.0	0.00	0.31	0.75	1.31	2.00
	α_2 = 0.2	-5.00	-4.69	-4.25	-3.69	-3.00
	α_2 = 0.4	-10.00	-9.69	-9.25	-8.69	-8.00
α_1 = 0.4						
	α_2 = 0.0	0.00	0.56	1.25	2.06	3.00
	α_2 = 0.2	-5.00	-4.44	-3.75	-2.94	-2.00
	α_2 = 0.4	-10.00	-9.44	-8.75	-7.94	-7.00

Note: Table elements give that average value of $\beta_2^2 + \beta_1\beta_2\alpha_1 - \alpha_2\beta_1^2$. Corresponds to Table 1 in the main text.

Table 1 in the main text shows the rejection rates for the proposed Wald statistic when the autoregressive parameter in the residual process ρ is set to 0.4. Table 2, below, shows the corresponding rejection rates when $\rho = 0.0$ and Table 3 shows the rejection rates when $\rho = 0.2$.

Table 2: Wald Test for ADL(2,1) against PA(1) with residual autocorrelation, ρ = 0.0

	β_2 =	0.00	0.25	0.50	0.75	1.00	0.00	0.25	0.50	0.75	1.00
		$T = 50$					$T = 100$				
α_1 = 0.0											
	α_2 = 0.0	0.06	0.06	0.07	0.13	0.26	0.05	0.05	0.09	0.22	0.48
	α_2 = 0.2	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
α_1 = 0.2											
	α_2 = 0.0	0.06	0.08	0.17	0.39	0.68	0.06	0.10	0.31	0.68	0.95
	α_2 = 0.2	1.00	1.00	1.00	0.98	0.91	1.00	1.00	1.00	1.00	1.00
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
α_1 = 0.4											
	α_2 = 0.0	0.05	0.12	0.35	0.71	0.93	0.05	0.21	0.65	0.95	1.00
	α_2 = 0.2	1.00	1.00	0.98	0.90	0.59	1.00	1.00	1.00	1.00	0.90
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$T = 200$					$T = 1000$				
α_1 = 0.0											
	α_2 = 0.0	0.05	0.05	0.12	0.36	0.78	0.06	0.07	0.37	0.93	1.00
	α_2 = 0.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
α_1 = 0.2											
	α_2 = 0.0	0.05	0.16	0.55	0.94	1.00	0.05	0.53	1.00	1.00	1.00
	α_2 = 0.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
α_1 = 0.4											
	α_2 = 0.0	0.05	0.35	0.91	1.00	1.00	0.05	0.93	1.00	1.00	1.00
	α_2 = 0.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: Rejection rates are computed via using 1,000 replications of the ADL(2,1) model $y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_1 x_t + \beta_2 x_{t-1} + u_t$ where $u_t = \rho u_{t-1} + e$ and $e \sim N(0,1)$. The parameter $\alpha_1 = \{0.00, 0.20, 0.40\}$, the parameter $\alpha_2 = \{0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50\}$, the parameter $\beta_1 = 5$, and the parameter $\beta = \{0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50\}$. The parameter $\alpha = 0$ in all conditions. The reported 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.

Table 3: Wald Test for ADL(2,1) against PA(1) with residual autocorrelation, ρ = 0.2

	β_2 =	0.00	0.25	0.50	0.75	1.00	0.00	0.25	0.50	0.75	1.00
		$T = 50$					$T = 100$				
α_1 = 0.0											
	α_2 = 0.0	0.06	0.07	0.08	0.13	0.25	0.05	0.05	0.08	0.21	0.47
	α_2 = 0.2	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
α_1 = 0.2											
	α_2 = 0.0	0.06	0.08	0.18	0.39	0.68	0.05	0.10	0.31	0.68	0.95
	α_2 = 0.2	1.00	1.00	1.00	0.98	0.92	1.00	1.00	1.00	1.00	1.00
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
α_1 = 0.4											
	α_2 = 0.0	0.06	0.13	0.36	0.72	0.94	0.06	0.22	0.64	0.96	1.00
	α_2 = 0.2	1.00	1.00	0.99	0.91	0.59	1.00	1.00	1.00	1.00	0.91
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$T = 200$					$T = 1000$				
α_1 = 0.0											
	α_2 = 0.0	0.06	0.06	0.12	0.34	0.75	0.06	0.06	0.33	0.91	1.00
	α_2 = 0.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
α_1 = 0.2											
	α_2 = 0.0	0.06	0.14	0.54	0.94	1.00	0.05	0.49	1.00	1.00	1.00
	α_2 = 0.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
α_1 = 0.4											
	α_2 = 0.0	0.05	0.34	0.92	1.00	1.00	0.06	0.92	1.00	1.00	1.00
	α_2 = 0.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	α_2 = 0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: Rejection rates are computed via using 1,000 replications of the ADL(2,1) model $y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_1 x_t + \beta_2 x_{t-1} + u_t$ where $u_t = \rho u_{t-1} + e$ and $e \sim N(0,1)$. The parameter $\alpha_1 = \{0.00, 0.20, 0.40\}$, the parameter $\alpha_2 = \{0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50\}$, the parameter $\beta_1 = 5$, and the parameter $\beta = \{0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50\}$. The parameter $\alpha = 0.2$ in all conditions. The reported 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.

References

Sargan, John D. 1964. "Wages and prices in the United Kingdom: a study in econometric methodology." *Econometric analysis for national economic planning* 16:25–54.

Wilkins, Arjun 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.