Long Run Confidence: Estimating Confidence Intervals when using Long Run Multipliers

Mark David Nieman* and David A.M. Peterson†

August 2024

Abstract

Social scientists are often interested in the long run relationship between two variables where the dependent variable has dynamic properties. Though methods for calculating the long run multiplier for an independent variable are well-developed, calculating confidence levels and assessing statistical significance is often difficult, especially when panel lengths are relatively short. We build on insights from Webb, Linn, and Lebo (2019), who introduce a bounds approach for evaluating these relations without making assumptions about the series exhibiting stationary or unit root processes. We propose a Bayesian framework that uses a bounded prior on the lagged dependent variable to constrain estimates for the dynamic relationship to the plausible range of values arising from either stationary or integrated series, and then taking draws of the posterior distribution to summarize the credible region. Doing so allows for recovering estimates of the long run multiplier and its uncertainty, even for short time series. After highlighting the advantages of adopting this framework via Monte Carlo experiments, we replicate several existing studies to show how our method clarifies long run relationships that were found inconclusive using existing techniques.

*Assistant Professor of Political Science, University of Toronto.

†Lucken Professor of Political Science, Iowa State University.

Applied time series researchers often face difficult choices. It can be difficult to determine the best, let alone correct, specification. In most cases, the first step in this process is to test the stationarity of our series and make our modeling choices based on these diagnostics. Philips (2018), for instance, provides a remarkably useful flowchart of choices that one should make based on the results of a set of diagnostics. The difficulty is that many of the tests are low powered and our series frequently have a small number of observations. Too often, different tests will provide inconclusive or contradictory results. The applied researcher, then, has to do what he or she thinks is best and hope that readers and reviewers agree.

For a short period, a misreading of De Boef and Keele (2008) led some scholars to act as if a generalized error correction model (GECM) was a panacea for these problems. Grant and Lebo (2016) noted the difficulties with this approach and reiterated the need for effective diagnostics of the properties of time series. Grant and Lebo (2016) and Philips (2018) serve as useful summaries of the issues that time series analysts face and introduce the associated bounds testing procedure of Pesaran, Shin, and Smith (2001) (PSS) to political science. This approach recognizes the uncertainty we often have about the stationarity of our independent variables. Unfortunately, these early discussions still treat the diagnostics about the dependent variable as definitive. Philips (2018) has clear proscriptions for how to approach modeling time series if the dependent variable is stationary or non-stationary. If one can trust the knife-edged tests of stationarity, then the recommended approach is relatively straightforward and one can simply follow the recipe that Philips provides.

Webb, Linn, and Lebo (2019) (WLL) reminded practitioners that these unit root tests are rarely certain. They advocate for a bounds approach that focuses on the long run multiplier (LRM) that summarizes the relationship between the independent and dependent variables. They show that the significance of the LRM is a test of the presence of the long run relationship between the variables, regardless of the stationarity. Importantly, they also provide the bounds of the t-test of the LRM that applied researchers can use to inform their conclusions about the significance of the relationship between the variables. It is difficult to overstate the importance of this result for applied time series researchers. One can use the test they provide regardless of the clarity of the stationarity tests for the dependent and independent variables. All that is needed is to estimate either a GECM or an autoregressive distributed lag model, calculate the LRM and an estimate of the uncertainty of the LRM, and calculate the ratio of these two.

This solution is straightforward and elegant. It has a single problem: the estimates of the uncertainty in the LRM are complicated. The LRM is a ratio of two coefficients and there is "no simple formula for calculating the standard error of a ratio of coefficients" (Webb, Linn, and Lebo, 2019: 287). There are two methods for approximating the variance of the LRM—the delta method and the Bewley transformation—and WLL show that the distribution of the ratio of the LRM and its standard error is not standard. Their solution is to run a series of dynamic simulations and develop critical values of the test statistic.

This is a smart approach, but it has one limitation. The bounds method that WLL use has a range of values where the hypothesis of a long run relationship between X and Y is rejected, a range where it is not rejected, and a range of values that is indeterminate. Their advice is to treat results that fall in this indeterminate range as failing to reject the null hypothesis of no relationship and be transparent about the lack of a definitive conclusion. This is likely to frustrate many applied researchers. An indeterminate answer to a research question is generally unsatisfying, even if it is intellectually honest.

This frustration is potentially unnecessary. In this manuscript, we develop a very simple Bayesian estimator of the LRM that does not have this indeterminacy. We start by using a bounded prior for the estimated coefficient on the lagged DV that constrains the resulting dynamic relationship to the plausible range of values from either stationary or integrated series. We then take advantage of the well-known property of Markov chain Monte Carlo (MCMC) models where one can estimate and summarize the distribution of functions of parameters (e.g., ratios of coefficients) directly from the posterior distribution (Gelfand et al, 1990; Murr, Traunmüller, and Gill, 2023). This framework requires minimal additional assumptions over the approach suggested by WLL and is easy to estimate in most software. One could incorporate more information through the use of informative priors in the estimation, but this is not our intention here. We show that very diffuse priors enable the use of MCMC methods and the direct estimation of uncertainty of the LRM.

In the next section of this paper, we revisit the results presented by WLL, demonstrating the importance of the significance tests of the LRM. Next, we provide the very simple MCMC approach to testing for the presence of an LRR. We use two Monte Carlo experiments to demonstrate key properties of our approach and compare them with the bounds approach. We then apply our approach to three empirical applications: two included in the original WLL work and one recent additional publication. These results show that the Bayesian approach to estimating the significance of the LRM and the presence of the LRR can lessen the uncertainty and indeterminacy that researchers face.

Long Run Relationships and Hypothesis Testing

Most applied time series work in political science is intended to test for some relationship between one or more weakly exogenous independent variables, X, and a dependent variable y.[1] The key to these models is the existence of an LRR between X and y, which implies that there is a long-run equilibrium between the two. While the presence of an equilibrium means that the variables tend to not change over time, the practical implication is that it is the place where the variables tend to return to when they do deviate (Banerjee et al, 1993; Box-Steffensmeier et al, 2014; Burke, Hunter, and Canepa, 2017; Webb, Linn, and Lebo, 2019, 2020).

The particular nature of the equilibrium depends on the stationarity of the series. For a stationary series, the mean is the equilibrium. It will eventually revert to it when it deviates from this mean. The particular type of equilibrium depends on the relationship

[1] We follow WLL and denote X as a set of multiple regressors and x to indicate a single regressor. We also note that there are many reasons why a researcher might be interested in multiple dependent variables, but we are focusing our attention on models with a single dependent variable.

between X and y. As WLL note, if the equilibrium of Y is a function of X, then there is a *conditional stationary equilibrium* and if the equilibrium of y does not depend on X, then there is a *unconditional stationary equilibrium*. In contrast, a variable that is non-stationary, by definition, does not have an equilibrium level that it will tend to return to. The notion of the "random walk" is that this type of series will move randomly and not tend to move back to some mean level. This type of series, however, can have an equilibrium based on a relationship with X. If X also has a unit root, then a cointegrating equilibrium can exist between X and y, where y will tend to move together over time. In this case, there is a *cointegrating equilibrium*.

Traditionally, diagnosing the type of equilibrium is an essential step in testing for the LRR between X and y. The tests we use for our hypotheses and the critical values of those tests depend on these diagnostics. Getting the diagnostic wrong likely means that we will get the substantive conclusions wrong. This is the heart of an exchange on time series analysis in *Political Analysis*. Grant and Lebo (2016) demonstrate that if the researcher gets the diagnostics incorrect, or if they simply run a GECM without paying attention to the properties of the series, they can make remarkable errors in their hypothesis tests.

But how was an applied researcher supposed to move forward? If we knew the type of equilibrium possible for our variables, then we would know which model to use. The flowchart provided by Philips (2018) provides clear guidance on this. If X and y are all stationary, run an autoregressive distributed lag (ARDL) model. If the bounds test suggests cointegration, estimate a GECM. If there is not enough evidence to conclude that there is cointegration, difference the variables and then run an ARDL. This advice is straightforward and helpful and the bounds approach created by PSS is an excellent step forward.

The problem with this approach is, as WLL note, that it starts with the assumption that one can definitely diagnose if the dependent variable is stationary. This is often much harder than it sounds. Unit root tests have low power, particularly when we are working with the short time series that are common in political science. It gets more complicated because we have to make choices about trend, drift, and serial correlation that will change the test. Given the large number of decisions and tests available, all too often researchers end up with conflicting evidence from their diagnostics about the nature of the series. As a result, they have to hope that the results are robust enough that the specification choice they make that they reach the same conclusion regardless of which approach they choose.

This is the motivation behind the work of WLL. They start with the error correction model (ECM) setup:

$$\Delta y_t = \alpha_0 + \alpha_1^*(y_{t-1} - \lambda x_{t-1}) + \beta_0^* \Delta x_t + \epsilon_t \tag{1}$$

where the LRM is represented as λ. This captures the total effect of a one-unit change in x on y summed over time. The y_{t-1} - λx_{t-1} piece of the equation is the long-run equilibrium relationship. The α_1^* term is the error correction that accounts for how fast the system returns to equilibrium after a shock. The actual estimation of this model is usually done via the GECM setup:

$$\Delta y_t = \alpha_0 + (\alpha_1 - 1)y_{t-1} + \beta_0 \Delta x_t + (\beta_0 + \beta_1)x_{t-1} + \epsilon_t \tag{2}$$

or as an ARDL model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t \tag{3}$$

Of particular interest is the LRM, which is calculated as $\frac{\beta_0+\beta_1}{\alpha_1-1}$ in the GECM, and as $\frac{\beta_0+\beta_1}{1-\alpha_1}$ for the ARDL. Which model is used is something of a matter of taste as they are mathematically equivalent to one another (De Boef and Keele, 2008: 189–190).

While recovering a point estimate of the long run relation is relatively straightforward by simply inputting the estimated coefficient into the appropriate formula, calculating uncertainty is more complicated. As De Boef and Keele (2008) note, neither the ARDL nor the GECM provide a direct estimate of the standard error of the LRM. Since the LRM is a ratio of coefficients, the calculation of the variance of the ratio of coefficients with known variances can be used. The formula is:

$$Var(\frac{a}{b}) = (\frac{1}{b^2})Var(a) + (\frac{a^2}{b^4})Var(b) - 2(\frac{a}{b^3})Cov(a,b). \tag{4}$$

There are two approaches used approximate the variance of this ratio and estimate uncertainty in the long run relationship. The first is to calculate of the LRM from an ECM and use the Bewley (1979) transformation, which estimates the variance of the LRM directly.The Bewley transformation is:

$$y_t = \alpha_0\phi - \alpha_1\phi\Delta y_t + \phi(\beta_0 + \beta_1)x_t + \phi\beta_1\Delta x_t + \phi\epsilon_t \tag{5}$$

where $\phi = (\frac{1}{\alpha_1 - 1})$ and an instrument for Δy_t is calculated as the predicted values from the equation $\Delta y_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 x_t + \gamma_3 \Delta x_t + \epsilon_t$. The LRM is the coefficient on x_t from equation 5. The second approach is to use the delta method. The delta method relies on expanding a random variable—in this case the LRM—via a Taylor series and calculating the resulting asymptotic variance of this estimate.

WLL demonstrate convincingly that a clear test of the presence on long run relationship between X_t and Y_t is captured by the significance of the LRM. As they note "*Thus, a nondegenerate, or valid, equilibrium relationship between y_t and x_t requires the LRM to be nonzero*" (pp. 286-287, emphasis in original). Moreover, they demonstrate that this is true regardless of the stationarity of y_t, helping resolve much of the uncertainty in pre-analysis specification tests of the time series. This is a vitally important result. Additionally, because the LRM is calculated separately for each of the independent variables, this approach allows the researcher to know which of the variables have a significant long run relationship with y_t, while the ECM-based tests only indicate that at least one independent variable has a long run relationship.

What WLL also make clear, however, is that the interpretation of the specific parameters in the model depends entirely on the univariate properties of the individual series. Without knowing, with certainty, if these series are stationary, the traditional tests are indeterminate.

Given the importance of the LRM for specification testing, WLL empirically explore the appropriate distribution of the test statistics for the LRM based on the Bewley transformation. While the exact amount of information in the uncertainty estimates and the critical values for the LRM depend on the sample size and the degree of autoregression in y_t and x_t, in general they find that the critical values do no follow a standard distribution. Instead, they estimate these critical values via a stochastic simulation to determine the bounds of the test. Their conclusion is that most empirical tests of the LRM are likely over confident in the hypothesis tests. More importantly, they develop the bounds for the hypothesis tests of the long run relationship.[2] These bounds will guide a researcher to conclude that whether or not there is a long run relationship.

This is a tremendous step forward for applied time series. What may be unsatisfying for many researchers, however, is that the bounds have a relatively large range of middle values that are inconclusive. For many empirical research questions, a researcher may end up with the unsatisfying result of a test statistic between the bounds and an uncertain conclusion.

So how should an applied researcher interpret indeterminate results, or results that are near the bounds? WLL de facto treat these cases as failing to reject the null hypothesis, whether or not a coefficient has cleared the lower bound and approaches the upper bound. Yet, in practice, a long run relationship may exist even if estimates may fail to significance owing to a number of reasons, ranging from low power, a lack of precision, or issues related to equation balance (Keele, Linn, and Webb, 2016: 34-35).

Low power and a lack of precision are related, but distinct, issues. Low power may stem from a short time series or data that are truncated on either their starting or ending points owing to how they were collected. A lack of precision may result from measurement error in

[2] Webb, Linn, and Lebo (2020) provide an expanded set of critical values used to denote upper and lower bounds, by both the number of observations [25, 50, 75, 150, 500, 1000] and α-level [.01, .05, .10].

one of the variables or from a time series simply being too short to return to equilibrium. In either case, when the number of periods in the time series T is small, for example, the estimate of α_1, the coefficient of the LRM, can be imprecise. In settings where dynamic processes are not in play, these issues simply result in inflated standard errors and not many other issues. When dynamic processes are relevant, however, the inflation of standard errors is amplified through the LRM and is more problematic. For instance, if the confidence interval of α_1 includes 1, the variance calculation can produce nonsensical results: if the value of α_1 has probability mass at 1, the LRM is undefined for that point, while if there is mass where $\alpha_1 > 1$, the denominator will be negative. In each case, the calculation of the variance breaks down.

Similar issues can complicate checks of equation balance. Equation balance refers to ensuring that X and y, individually or collectively, have the same order of integration. Pickup (2022) identifies common sources of confusion around balance when using time series data and illustrates the effect this has on one's inferences. Pickup and Kellstedt (2022) provide several procedures for implementing both theoretical and empirical checks of equation balance. Yet, even if a X and y meet the theoretical checks, empirical checks for determining the order of integration and cointegration is difficult and can result in contradictory results, especially for short time series (Pickup and Kellstedt, 2022: 301).

Given these various issues, which are all too common across the social sciences, applied researchers face numerous practical difficulties. Even if they meet theoretical standards of equation balance, their data may simply make it difficult to meet decisively demonstrate that it is stationary. Thus, treating any result that fails meet the more stringent upper bound criteria of the bounds test as the threshold for identifying a long run relationship, whether the lower bound was exceeded or not, may be impractical for advancing our understanding of relationships for which the data simply do not play nice and clearly meet (or violate) our tests' stationarity assumptions.

A Bayesian Approach

Our approach to this problem is to adopt a Bayesian framework. This allows us to exploit the limits on the range of plausible values of the coefficient for the lags of the dependent variable by adding a prior on the coefficient for the lagged DV that will constrain it to be strictly between -2 and 0 in the GECM and -1 and 1 in the ARDL.[3] One could specify a diffuse prior that places equal probability on all values between these bounds. Alternatively, a more informed prior, which places more weight to values closer to 0 or either bound, may be practical in applied settings where one can incorporate existing knowledge.[4]

One way to think about this prior is that it is simply the formalization of plausible range of values on the dynamic relationship between the dependent variable and its lags that researchers are already making when they estimate a model like that in Equation 2 or 3. When a researcher treats a series as stationary, she is assuming that the root of the characteristic equation of the time series is less than one. When the series is integrated, the root of the characteristic equation is exactly one. The use of this prior constrains the estimate of α_1 to be no greater than 1, precisely the implication of treating the dependent variable as stationary or integrated.

An additional benefit of a Bayesian approach is that it works extremely well with samples with a small T (McNeish, 2016). Such samples are frequently the types of time series that applied researchers use and that are the cause of so many diagnostic problems.[5] There are many cases where assuming stationarity is appropriate for small samples. The small T on its own suggests that the stationarity tests will often be weak, placing theoretical and substantive knowledge of the time series at a premium. Undoubtedly, if a coefficient on the

[3]The GECM in Eq 2 assumes that $-2 < \alpha_1 < 0$ and the ARDL in Eq 3 assumes that $-1 < \alpha_1 < 1$. See Keele, Linn, and Webb (2016: Table 1) for a summary of the error correction rates and long run equilibria.

[4]If one is estimating an ARDL, for example, and has theoretical reasons to expect the coefficient on the lagged DV to approach 1, then a prior such as $\mathcal{B}(5, 2)$, where the pmf is massed near 1, could be used. If, instead, a researcher believes that the coefficient on the lagged DV is positive, but no other information on its pmf, a prior such that $\alpha_1 \sim \mathcal{B}(1, 1)$ would apply a uniform distribution between 0 and 1.

[5]An alternative strategy for estimating time series with small T is to use a transformed-likelihood estimator, such as an orthogonal reparameterization estimator (Pickup and Hopkins, 2022).

lagged dependent variable is not significantly different than 1, that series is unlikely to reject the null of a unit root in stationarity tests. Researchers, however, often have more knowledge about the properties of their dependent variable beyond the small set of observations used. Researchers who are limited by the time-span of their independent variables, for instance, may have a series that is stationary with larger T than they can use in their analyses. In other cases, time series of comparable data, whose properties are better known, may exist and offer a researcher additional contextual information.[6]

The actual estimation of the model is carried out via MCMCs. This allows us to calculate the distribution of the posterior for all of the coefficients directly. Rather than using an asymptotic equivalent to the confidence interval of the LRM or relying on a formula that is not easily available in most statistical output, we can calculate the LRM for each of the draws from the posterior in the MCMC and use this distribution to summarize the credible region of the LRM. This is one of the virtues of inference from the posterior of an MCMC: a researcher can estimate the distribution of functions (such as ratios) of an unknown parameter or parameters directly from the posterior distribution of the MCMC (Gelfand et al, 1990; Murr, Traunmüller, and Gill, 2023).[7] As such, we do not need to rely on the asymptotic properties of the variance estimator of the LRM and should provide more accurate estimates of the credible region for the LRM than either the Bewley method or the delta method (especially for small T).

Finally, using a bounded prior keeps estimates of uncertainty of the LRM firmly within their theoretical bounds. If the confidence interval for the LRM reaches 1 for either the GECM or ARDL, then denominator of that LRM will include zero. In that case, the estimation of the variance of the LRM will be "mildly explosive"[8] (which is bad). The prior thus keeps estimates of uncertainty within the same theoretical bounds as the point estimate,

[6]For instance, information garnered from existing data or studies using comparable administrative data from within a country, region, or similar countries are often available for researchers of American politics (other US states/counties), comparative politics (provinces, nearby countries), or international relations (countries with similar economic development).

[7]This approach can be easily implemented in common statistical programs, such as R or Stata.

[8]We borrow this phrase from Hill and Peng (2014, 293) and Hill, Li, and Peng (2016, 126).

providing more substantively plausible and theoretically-informed results.

To demonstrate the usefulness of adopting a Bayesian approach for estimating the LRM, we conduct two Monte Carlo experiments. The first experiment compares point estimates and coverage estimates for the LRM—using the bounds approach from a Bewley transformation of the ECM, and the Bayesian approach—under varying levels of univariate autocorrelation of x and y and different time series lengths, when there is no long run relationship. These conditions allow for assessing well each estimation approach separates a spurious long run bivariate relationship from actual univariate dynamics.

The second experiment illustrates how to apply the Bayesian approach in a more realistic (but still controlled) setting where the researcher is interested on testing and reporting the instantaneous and long-run effects of an independent variable. In this experiment, we report point and uncertainty estimates for x, a lagged y, and the LRM, where there is a moderate long run relationship between x and y as well as moderate autocorrelation between x and its lag. This specification allows us to evaluate how well each approach captures a true long run relationship under conditions that a research would experience in practice.

Following these experiments, we compare the results and provide a general discussion for applied researchers, focusing on trade-offs and gains from each approach.

Monte Carlo Experiment #1

For the first Monte Carlo experiment, we generate our data replicating the dynamic simulation process used by WLL to identify bounds for the LRM. We begin by generating two independent autoregressive processes, such that $y_t = \rho_y y_{t-1} + \epsilon_y$ and $x_t = \rho_x x_{t-1} + \epsilon_x$, with the errors drawn from separate standard normal distributions and $T = 75$. While the error terms are each stationary, the values of ρ_y and ρ_x are set to be either 0 or 1—reflecting $I(0)$ and $I(1)$ for each variable. This gives four scenarios: one where $\rho_y = 0$ and $\rho_x = 0$, a second with $\rho_y = 0$ and $\rho_x = 1$, another where $\rho_y = 1$ and $\rho_x = 0$, and finally a case with $\rho_y = 1$ and $\rho_x = 1$. In the true data generating process there is no long run relationship between

y_t and x_t; therefore, any (mis)identified relationship is strictly due to the dynamics induced through the univariate autoregressive processes.

Using these data, we continue to follow WLL by estimating the LRM and its uncertainty using the Bewley tranformation from Eq 5. We then also estimate a Bayesian ECM. As described above, one of the advantages of the Bayesian ECM is that uncertainty for each parameter, including those of functions, can be directly estimates from the posterior distribution of the MCMCs. That is, rather than an approximation for the uncertainly for the LRM that relies on asymptotic properties, as is the case with the Bewley transformation, the direct estimates from the Bayesian estimator should be more accurate, particularly when T is relatively short and asymptotic properties are least likely to hold.

The Bayesian ECM is specified as in Eq 2, with diffuse priors of $\mathcal{N}(0, 20)$ for the constant and the coefficients associated with Δx_t and x_{t-1}, an uninformed prior of $\mathcal{U}(0, 2)$ on the coefficient for lagged y, and the prior for the variance distributed $\mathcal{G}(1, 10)$. Recall that the prior on the coefficient on y_{t-1} for an ECM formalizes the specification of the dynamic relationship between the dependent variable and its lagged values and prevents it from taking explosive values that do not return to the LRR equilibrium. Each Bayesian ECM is estimated using 5,000 MCMCs after a 2,500 burnin and thinning of 10.

For each combination of ρ_y, ρ_x from the data generating process, we estimate the LRM from the Bayesian and Bewley specification for two lengths of T, such that $T \in \{25, 75\}$.[9] The varying lengths for T allow us to look at both how well the estimates fair for short and moderately short time series that are common to social science data.

Table 1 reports summaries for estimates of the LRM from 20,000 simulations under each of the 8 scenarios (four ρ_y, ρ_x possibilities times two lengths of T).[10] For both the Bayesian and Bewley estimates of the LRM, we report a point estimate (posterior median for the Bayesian estimator, average point estimate for the Bewley estimator) and its coverage rate.

[9]Both lengths of T are taken from the same original time series: when $T = 25$, the first 25 observations are used; when $T = 75$, all of the observations are used.

[10]To clarify, within each individual simulation of the first Monte Carlo experiment, the Bayesian ECM estimates are summaries based on its own 5,000 MCMCs following a 2,500 burnin. Wheels within wheels.

The point estimate allows for assessing bias in the estimate of the LRM, while the coverage rates offer insights into how well the different estimation strategies perform in returning accurate estimates under varying conditions that are common in applied work (Hopkins et al, 2023). Lower coverage rates would suggest that, even if an estimator is unbiased on average, its results are less unreliable in any particular application. Coverage rates for the Bayesian ECM report how frequently the true value is within the estimate 95 percent credible intervals. For the Bewley model, we report the coverage rate using the bounds approach suggested by WLL; that is, we construct 95 percent confidence intervals for both the lower and upper bound at each length of T, using the appropriate t-statistics identified by Webb, Linn, and Lebo (2020: Table 2).[11] We also report the percent of indeterminate cases arising from incongruent outcomes between using lower and upper bound.

There are several notable results from the Monte Carlo experiment. In terms of bias, estimates of the LRM are similar for both the Bayesian and Bewley approaches. When $\rho_y = 0$, the Bayesian and Bewley models each correctly recover estimates showing no long run relationship. When $\rho_y = 1$, both show limited bias, though a quick glance of their coverage rates indicates that, even in these cases, the overwhelming majority of the time their credible and confidence intervals correctly contain zero. That both estimators perform better when $\rho_y = 0$ is unsurprising, and that each approach correctly predicts the (lack of a) long run relationship provides initial confidence in the average accuracy of either approach.

Next, we focus on coverage rates. First, the Bayesian ECM recovers the true value in the overwhelming majority of cases, with rates of approximately 95% when $\rho_y = 0$ and a slightly lower 88% when $\rho_y = 1$. When using the bounds approach with the the Bewley estimates, the upper bound recovers the true value in nearly all simulations, with only the scenario of $\rho_y = 1, \rho_x = 1$ being at 95% (which coincides, of course, the aim of the WLL's bounds approach). The lower bound, however, performs much more poorly, with a high coverage rate of 78% and a low of 58%. The indeterminate range, where applied researchers cannot

[11] In our experiment, with one X variable and $T = 25$, the t-statistic for the lower bound is 1.25 and for the upper bound is 3.79. When $T = 75$, the corresponding t-statistics are 1.06 and 3.68, respectively.

Table 1: LRM Estimates with Varying Autocorrelations and Sample Size.

	$T = 25$			
	$\rho_y = 0, \rho_x = 0$	$\rho_y = 0, \rho_x = 1$	$\rho_y = 1, \rho_x = 0$	$\rho_y = 1, \rho_x = 1$
Bayesian ECM				
Median	-0.001	-0.001	0.002	-0.001
Coverage	0.961	0.958	0.888	0.902
Bewley Transformation				
Mean	0.001	-0.001	-0.012	0.091
Lower Bound Coverage	0.763	0.749	0.780	0.648
Upper Bound Coverage	0.999	0.997	0.998	0.950
Indeterminate Range	0.235	0.248	0.218	0.303
	$T = 75$			
	$\rho_y = 0, \rho_x = 0$	$\rho_y = 0, \rho_x = 1$	$\rho_y = 1, \rho_x = 0$	$\rho_y = 1, \rho_x = 1$
Bayesian ECM				
Median	0.000	-0.000	-0.000	-0.003
Coverage	0.948	0.948	0.878	0.882
Bewley Transformation				
Mean	-0.000	-0.000	-0.002	-0.248
Lower Bound Coverage	0.702	0.697	0.763	0.583
Upper Bound Coverage	0.999	0.999	0.999	0.951
Indeterminate Range	0.297	0.302	0.237	0.369

Note: Bayesian estimates are median and 95% credible intervals from the posterior distribution. The LRM (and uncertainty) for the traditional ECM is estimated using the Bewley transformation. The coverage range for the Bewley transformed LRM is calculated using a t-statistic of 1.25 for the lower bound and 3.79 for the upper bound when $T = 25$, a t-statistic of 1.06 for the lower bound and 3.73 for the upper bound when $T = 75$ (Webb, Linn, and Lebo, 2019, 2020).

confidently reject the null hypothesis nor fail to reject it, is never less than 21% and reaches nearly 37% in one scenario.

Second, when looking *within* each scenario of ρ_y, ρ_x across each T, the Bayesian ECM returns similar coverage rates. This reflects the fact, of course, that Bayesian statistics does not rely on large sample sizes (i.e. the Central Limit Theorem) for calculating results (see, e.g., Miočević, Levy, and van de Schoot, 2020: 1). In contrast, the Bewley coverage for the upper bound remain consistent, whereas that of the lower bound actually decrease as T increases. This stems from a smaller t-statistic being applied, when using the bounds approach, as the number of observations increase. As a result, the proportion of the indeterminate range between the lower and upper bounds increases with the size of the sample.

Third, the percent of cases in the indeterminate range for the bounds approach based on the Bewley estimates are greatest when $\rho_x = 1$. This result is substantively meaningful, as many covariates used as control (or even independent variables) included in panel data either do not change, or do not change very much, over time. For instance, it is well known in international relations that a country's regime type is usually relatively stable for long periods of time. Even in cases where these covariates do change, such as the GDP of a state, province, or country, they are often primarily a function of the own prior value. Each of these are cases where ρ_x would approach 1. This is also something, of course, that can be evaluated and known prior to estimating a dynamic model.

Fourth, the Bayesian ECM provides the greatest coverage when $\rho_y = 0$, with a decrease of 5 to 7 percentage points when $\rho_y = 1$. This decrease in accuracy holds regardless of the length of the time series. This result may, however, reflect our use of a diffuse prior, which gives equal weight to all theoretically possible values of the lagged y, thus pulling it towards the lower and upper bounds. Conversely, the Bewley coverage rates are highest when $\rho_y = 1, \rho_x = 0$, even outperforming itself compared to when y was not dynamic. However, the Bewley approach performs its worst when $\rho_y = 1, \rho_x = 1$, returning coverage rates below 65% for the lower bound.

These results highlight key small sample properties of LRM estimates from a Bayesian ECM and applying the bounds approach with Bewley estimates under varying univariate dynamics when the true LRR is zero. While both are generally unbiased, coverage rates are impacted by the univariate dynamics. When y behaves nicely, then the benefits of the Bayesian approach are most evident: high coverage rates without the inconvenience of wide indeterminate ranges. When y is less well mannered, the coverage rates for the Bayesian approach drop slightly[12] while the indeterminate range for the bounds approach remain large.

Monte Carlo Experiment #2

For the second Monte Carlo experiment, we generate a simple dynamic model with a moderate long run relationship between X and y, and mimic common features of real world

[12]These coverage rates, of course, would improve if a more informed prior were used.

data by inducing a mild autoregressive process between X and its lag. More specifically, we generate the endogenous variable so that $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$ where $\alpha_0 = 0$, $\alpha_1 = 0.5$, $\beta_0 = 0.5$, and $\beta_1 = 0.25$. We generate the exogenous variable so that $X_t = \gamma X_{t-1} + \eta_t$ where $\gamma = 0.5$. Both η_t and ϵ_t are drawn from a standard normal where $cov(\eta_t, \epsilon_t = 0)$. The LRM, given this specification, is $\frac{0.5+0.25}{1-0.5} = 1.5$. Since we are interested in the small sample properties of each approach, we set $T = 25$. This experiment illustrates how well each approach is able to help researchers make correct inferences when there is a LRR between their variables of interest.

We estimate both a Bayesian ECM and a traditional ECM to recover parameter estimates, specified as in Eq 2. The LRM is estimated directly from the posterior distribution in the case of the Bayesian ECM, and via the Bewley transformation for the traditional ECM. We give the Bayesian ECM diffuse priors of $\mathcal{N}(0, 20)$ for the constant and the coefficients associated with x_{t-1} and Δx, a semi-informed prior of $\mathcal{U}(-1, 0)$ on the coefficient for y_{t-1} such that the LRR cannot be negative (e.g., assumed to either be positive or have no effect), and a diffuse prior of $\mathcal{G}(1, 10)$ for the variance. Each individual Bayesian ECM is estimated from 5,000 MCMCs after a 2,500 burnin with a thinning of 10. Finally, the Bewley estimate of the LRM for the tradition ECM is specified as in Eq 5.

Table 2 reports summaries for each estimation approach based on 20,000 simulations. The first column reports the true value of the ECM estimates based on the data generating process described above;[13] the second column reports the median value, and the accompanying 95 percent credible interval, taken from the posterior distribution for the Bayesian ECM; and the last column reports the mean coefficient and standard error for the traditional ECM. Below the parameter for the instantaneous effects is the estimated LRM and its uncertainty; for the traditional ECM, these are obtained via the Bewley transformation. Finally, the bottom of the table reports the coverage rates from each estimator, with the lower and upper bounds based on the t-statistics identified by WLL used for the Bewley estimates, along with the

[13]Recall that the ADL and ECM specifications are mathematical equivalents of one another (De Boef and Keele, 2008; Keele, Linn, and Webb, 2016).

Table 2: Error Correction Model Parameter Estimates with a Moderate LRR and Univariate Autocorrelations, and $T = 25$.

	True Value	Bayesian ECM	ECM w/ Bewley Trans. & Bounds
Y_{t-1}	-0.5	-0.577	-0.594
		[-0.88, -0.25]	(0.172)
X_{t-1}	0.75	0.797	0.815
		[0.220, 1.368]	(0.273)
ΔX	0.5	0.494	0.495
		[0.011, 0.976]	(0.224)
Constant	0	-0.003	-0.003
		[-0.512, 0.506]	(0.239)
LRM	1.5	1.452	1.454
		[0.173, 3.469]	(0.543)
Bayesian Coverage Rate		0.683	
Lower Bound Coverage			0.940
Upper Bound Coverage			0.292
Percent in indeterminate Range			0.648

Note: Bayesian estimates are median and 95% credible intervals from the posterior distribution. The LRM (and uncertainty) for the traditional ECM is estimated using the Bewley transformation. The coverage range for the Bewley transformed LRM is calculated using a t-statistic of 1.25 for the lower bound and 3.79 for the upper bound (Webb, Linn, and Lebo, 2019, 2020).

range of indeterminate values where the lower and upper bounds give conflicting inferences.

Starting with the estimates associated with the specified variables, both the Bayesian and traditional ECM return similar values in terms of bias and produce similar inferences in determining which factors are determinants of Y, with all measures of uncertainty including the true value, on average. The same holds for the estimate of the LRM, with each finding a positive, non-zero effect, on average.

Turning to the coverage rates, however, there is greater divergence. While the entirety of the Bayesian estimator's 95 percent credible interval range is greater than zero in 68.3% of the simulations, the 95 percent confidence intervals estimated using the bounds approach offers for a more muddled conclusion.[14] Though the lower bound criteria returns estimates that are greater than zero for the entirety of the 95 percent confidence interval in 94% of the

[14] We report the coverage rates for the entirety of the credible range in order to make a more straightforward comparison between estimators in the current Monte Carlo experiment. In practice, however, one could calculate and report the percent of individual MCMC draws above/below zero to provide more precise confidence in the probability of a LRR between X and y. We demonstrate this in the next section.

simulations, the more challenging upper bound criteria does so in only 29.2% of the time. The result is that, in almost two-thirds of the simulations (64.8%), the bounds approach offers conflicting guidance on whether a LRR exists between X and y; this is despite the fact that a true LRR between X and y does exist by the construction of the experiment. Moreover, this type of scenario is where the bounds approach is most likely to be implemented by an applied researcher, since the dynamics of X and y are neither $I(0)$ or $I(1)$, but in-between.

There are several takeaways from the two experiments. One is that the bounds approach—especially applying the stricter criteria of using only the upper bound—reduces the risk of a type 1 error. Conversely, this risk is slightly higher, under some conditions, when using the Bayesian approach. The cost of reducing this risk, however, is that the more stringent upper bound threshold dramatically increases the risk of a type 2 error. The Bayesian approach, on the other, is much better able to correctly recover moderate LRR, even when the sample size is small. Another key finding is that, for the bounds approach, the size of the indeterminate range can be quite large, especially when sample sizes are small. This holds regardless of whether an actual LRR between X and y exists. This characteristic is likely to be especially unsatisfying for applied researchers.

So what is an applied researcher to do? Given the insights from our Monte Carlo experiments, we think that the benefits of adopting a Bayesian approach is fairly strong, especially when the sample size is relatively small. At the small cost of a very slightly lower coverage rate when the LRR is null and the autocorrelation in y is very high—the latter a condition that would be evident from pre-testing and manageable with even a slighly-informed prior—one gains far more precision in making theoretical inferences.

Applications

In this section, we replicate three existing time series papers using our Bayesian approach. The first two are the same as in the original WLL paper, replicating the work of Ferguson, Kellstedt, and Linn (2013) on the effect of policy on public mood and Lebo and O'Green (2011) on presidential success in Congress. The third application is a recent paper on aggregate levels of interest in politics in America by Peterson et al (2022). These three applications demonstrate that the Bayesian approach can resolve some of the indeterminate results in existing work.

The first application is a test of how policy mood responds to public policy. Mood, in this context, is the electorate's preference for the size of government. When the public wants the federal government to expand, mood will be higher. When the electorate is more conservative, favoring a smaller government, mood will be lower (Stimson, 2018). The actual measure of policy mood is based on Stimson's (2018) dyad ratio algorithm that combines thousands of individual survey questions to develop a single time series of the preferences of the American electorate. It is one of the foundational concepts in the study of macro politics in America. Durr (1993) was the first to argue that mood should respond to changes in the economy and the actual size of government. As the economy expands, Durr contended, the public would be likely to express support for a more expansive government because the stronger economy would make it easier to pay for the policies. When the economy contracts and finances become tighter, the public is likely to prefer lower taxes, leading mood to decrease. The second main antecedent of policy mood, public policy itself, has been a more robust predictor of changes in mood. This thermostatic model of public opinion holds that when policy moves in one direction, the public responds by moving in the opposite direction (Wlezien, 1995).

WLL chose the Ferguson, Kellstedt, and Linn (2013) paper as an application because the time series properties of policy mood are notoriously difficult to diagnose. As WLL note, several tests of the stationarity of policy mood give conflicting and ultimately confusing results. Thus, they continue, scholars have not reached a consensus on the stationary of mood. Across the literature, it seems like applied researchers have used almost every possible specification of mood's dynamic properties and the tests are so inconsistent that no one really knows the correct specification. This is precisely the type of inconclusive results that bedevil applied time series researchers. They are usually forced to make a decision about the stationarity of policy mood and treat it as true. The approach advocated by WLL allows researchers to directly incorporate the uncertainty of the specification tests into the hypothesis testing.

WLL use the GECM setup for their model of policy mood. Their specification is:

$$\Delta Y_t = \alpha_0 + (\alpha_1 - 1)Y_{t-1} + \beta_0 \Delta X_{1t} + (\beta_0 + \beta_1)X_{1t-1} + \beta_2 \Delta X_{2t} + (\beta_2 + \beta_3)X_{2t-1} + \beta_4 \Delta X_{3t} + (\beta_4 + \beta_5)X_{3t-1} + \beta_6 v_t + \epsilon_t \qquad (6)$$

where Y_t is policy mood at time t, x_1 is inflation, x_2 is unemployment, x_3 is the liberalness of existing policy, and v_t is an intervention for the Vietnam war. The conclusions that WLL reach are that the evidence is insufficient to support either the theory that policy mood responds to the economy or the thermostatic model of opinion. The effect of unemployment is unambiguously insignificant. The other two LRMs, however, are less clear. The t-statistics are between the upper and lower bounds. While these are technically indeterminate results, WLL suggest that the researcher should fail to reject the null hypothesis with this pattern of results unless the researcher is confident that mood is a stationary series. Again, the stationarity tests are so inconsistent that this does not seem like a warranted conclusion.

Using our Bayesian approach to the model, we reach different substantive conclusions. We report the result in Table 3. The coefficients reported are medians from the posterior of the MCMC. We had four chains, each with a 25000 iteration burnin, followed by 5,000 iterations and a thinning of 5. The results, then, are based on a total of 4,000 draws from the posterior. The numbers in the parenthesis underneath are the 95 percent credible interval for the coefficients and the LRM from the posterior distribution. The point estimates of the coefficients are almost identical to the results in WLL using OLS. The main difference in the substantive conclusion is the significance of the LRM for the policy measure. In WLL, this

Table 3: A Model of Domestic Policy Mood: Second Quarter 1968 through Fourth Quarter 2010.

Variable	X_{t-1}	ΔX_t	LRM
Mood	-0.23		
	(-0.33, -0.13)		
Inflation	-0.12	-0.11	-0.51
	(-0.25, 0.01)	(-0.51, 0.29)	(-1.07, 0.06)
Unemployment	-0.08	0.93	-0.36
	(-0.28, 0.10)	(-0.02, 1.89)	(-1.49, 0.57)
Policy	-0.10	-0.17	-0.45
	(-0.20, -0.01)	(-0.58, 0.22)	(-0.95, -0.06)
Vietnam	1.82		
	(0.53, 3.11)		
Constant	19.57		
	(11.51, 27.83)		
T	169		

fell between the bounds, leading to an indeterminate conclusion. In our Bayesian approach, the 95 percent credible interval excludes zero, suggesting that there is a significant long-run relationship between public policy and public mood. In fact, in 98.7 percent of the the 4000 draws from the posterior the estimated LRM for the policy measure is negative.

For our second application, we replicate the work of Lebo and O'Green (2011) that explores the predictors of presidential success in Congress. The dependent variable in this model is the percentage of times the president wins a vote in the U.S. House by year from 1953 to 2006. The basic question in the research is if presidential approval gives the president the ability to persuade members of Congress to vote along with the president's policy preferences. While there is a long history of work predicting that approval gave the president more policy leverage, Edwards (2009) argues that the institutional features of the time determine how successful the president will be. Instead of being able to persuade members of Congress to vote counter to their predispositions, presidential success is predetermined by the preferences of members of Congress themselves. When presidents appear successful, it is really due to the partisan balance of Congress.

Table 4: A Model of Presidential Success, 1953–2006.

Variable	X_{t-1}	ΔX_t	LRM
Presidential success	-0.58		
	(-0.83, -0.33)		
Conditional party government	7.50	11.14	12.96
	(1.81, 13.36)	(5.54, 16.74)	(3.89, 22.51)
President's party House share	1.35	1.96	2.32
	(0.56, 2.14)	(1.41, 2.49)	(1.32, 3.29)
Presidential approval	0.09	0.30	0.15
	(-0.27, 0.43)	(-0.07, 0.67)	(-0.59, 0.73)
Constant	-34.64		
	(-73.05, 3.34)		
T	52		

The empirical application has three independent variables predicting presidential success. The main variable, and the key test of the presidential persuasion argument, is presidential approval. The other two variables capture the institutional balance of the U.S. House. The first is the percentage of House seats held by the president's party. The second is the conditional party government index (CPG) (Aldrich, Berger, and Rohde, 2002). WLL note that in these series, the unit root tests are ambiguous, making this work an apt choice for their bounds approach. The specification is also a straightforward ECM model.

In WLL's paper, they find a robust relationship between the share of the House held by the president's party and no evidence of a link between presidential approval and presidential success. The t-statistic capturing the relationship between the conditional party government measure and success, however, fell between the bounds. Again, WLL conclude that there is not enough evidence to support the hypothesis that conditional party government predicts presidential success.

We estimate our model with the same MCMC specifications as the previous application. Our results are similar to those of WLL for two of these three variables. The point estimates of the coefficients and the LRMs reported in Table 4 are almost identical to the ones reported in WLL. The credible region for the LRMs also leads us to conclude that there is a relationship between the share of seats held by the president's party and the president's success. We also find no evidence of a relationship between approval and presidential success. Our results differ on the effect of the CPG measure. The 95 percent credible region for the posterior estimate of the LRM excludes zero, indicating that there is a robust long-run relationship between the variables. With the MCMC approach, we have 4,000 draws from the posterior and can estimate the LRM for each of these iterations. For this variable, 99.3 percent of the draws from the posterior distribution are greater than zero.

Our final application explores the relationship between trust in government and interest in politics. Peterson et al (2022) argue that trust and interest trade-off in the electorate. When the electorate trusts the government more, the incentive to pay attention to politics lessens. Voters make a choice about whether or not to follow politics. For some, this is simply a habit or a hobby. For many people, however, it is something of a chore. They expect that when the electorate trusts government to look out for the public interest, the incentive to actively monitor the government lessens. Some of the electorate chooses to spend that time on things they find more enjoyable than politics. In contrast, went the electorate believes that the government is untrustworthy, the need for monitoring increases. The electorate will need to hold government officials more accountable. Peterson et al (2022) argue that this creates a tradeoff of normative evaluations of government. In general, higher levels of trust and higher levels of interest are believed to be markers of a stronger democracy. With this tradeoff, however, it is difficult to have high levels of both.

They rely on existing macro-level measures of trust in government and use the technique developed by Stimson (2018) to construct a new aggregate measure of macrointerest. Peterson et al (2022) report that both of these macro series are integrated and that the Engle-Granger two-step method supports the hypothesis that they are cointegrated. Based on these results, they estimate the relationship as an ECM and find support for their theory. They also develop two other alternative hypotheses about the effect of presidential approval and consumer sentiment on macrointerest and find no evidence for either relationship. They do find that presidential campaigns heighten levels of political interest. Finally, they argue that scandals and other major events may cause Americans to become more interested in politics and include a long list of major events during their timeframe. September 11 is the only of these events to have a significant effect on macrointerest.

The conclusion about the effect of trust on interest, however, is dependent on these specification tests and ignores the uncertainty in the stationarity and cointegration tests. If they are correct about the specification, then the evidence supports their theory. In the article, they report the confidence interval for their estimate of the LRM based on the Delta method, but not the t-statistic. Replicating their published work, we find that the t-statistic for the LRM is -2.71, which lies between the bounds provided by WLL and the result should be seen as inconclusive. Based on the bounds method, then, the main conclusion of Peterson et al (2022) is not clearly determined. They do find a significant short-term effect of changes in trust on changes in interest, but the LRM is not significant if one is not certain about the assumption that both trust and interest have unit roots.

To replicate the Peterson et al (2022) study using the Bayesian approach, we again use the same specification of the MCMC model as before. The results, reported in Table 5, have the same pattern of significant results as the original article. There is a negative relationship between trust and interest in both the long run relationship and the short term effect of changes in trust in government. Central for our application, the 95 percent credible interval of the LRM excludes zero, suggesting that there is a long-run relationship between the two. In fact, the estimate of the LRM is negative in 98.5 percent of the draws from the posterior. As a result, our analysis supports the main conclusion of the original article. We also replicate the results for the effect of September 11 and for presidential campaigns on macrointerest. Lastly, and consistent with Peterson et al (2022), we find no evidence that either ICS or presidential approval are linked to macrointerest.

Table 5: A Model of Macrointerest, 1973–2014.

Variable	X_{t-1}	ΔX_t	LRM
Interest	-0.14		
	(-0.22, -0.05)		
Trust	-0.09	-0.21	-0.62
	(-0.16, -0.01)	(-0.37, -0.04)	(-1.29, -0.13)
Consumer sentiment	0.00	0.00	0.01
	(-0.02, 0.02)	(-0.04, 0.04)	(-0.17, 0.16)
Presidential approval	0.01	0.00	0.04
	(-0.02, 0.03)	(-0.03, 0.03)	(-0.14, 0.28)
Presidential campaign	0.40		
	(0.23, 0.57)		
Watergate	0.55		
	(-0.66, 1.73)		
ABSCAM	1.52		
	(-0.96, 4.04)		
Jim Wright	-0.54		
	(-3.06, 2.05)		
Keating five	-0.23		
	(-2.72, 2.42)		
Clinton Impeachment	-0.29		
	(-2.13, 1.52)		
September 11	1.88		
	(0.01, 3.72)		
Hurricane Katrina	-0.11		
	(-2.50, 2.37)		
Invasion of Panama	-0.01		
	(-2.54, 2.45)		
Second Iraq War	-0.06		
	(-2.49, 2.38)		
Persian Gulf War	-0.96		
	(-3.48, 1.5)		
Constant	12.65		
	(4.77, 20.53)		
T	167		

Conclusion

Applied time series work can be frustrating for researchers. The need to get the dynamic properties of the series correct can bedevil a project. The weak tests for stationarity and the often inconclusive results from differing tests make time series analysis more complicated than might be appreciated. We have the hunch that there are likely numerous studies that have been started and stopped, and eventually shoved in a digital file drawer, because the researcher cannot be confident about the stationarity of the series they are working with. That is, time series researchers may not only be stymied from the same null results problem that we all face, but the uncertainty about which specification is the correct one can lead some to simply throw up their hands and give up on a project.

To this end, WLL provide a tremendous service by incorporating the uncertainty of the specification tests into their models and calculating a very clear set of bounds for when to conclude that there is an LRR between two variables. For some projects, this will be enough. If the results are clearly outside the bounds, the researchers know exactly what to do. But for many applied time series studies, the bounds approach will lead to inconclusive results. The necessity of the indeterminate zone of results, while intellectually honest, is likely to be unsatisfying for some. In this paper, we show that a Bayesian approach that directly estimates the LRM from the posterior of the model is one way to address the indeterminate zone of results. Using uninformative priors on most parameters, but directly incorporating the limits on the dynamic parameters in the model, we get better coverage in our Monte Carlos. This effect seems to be particularly pronounced for the small T types of time series that are common with social science data.

References

Aldrich, John H., Mark Berger, and David Rohde. 2002. The Historical Variability in Conditional Party Government, 1874–1944. In *Party, Process, and Political Change in Congress*, eds Brady, David W., and Mathew Daniel McCubbins. Stanford University Press. 17–35.

Banerjee, Anindya, Juan Dolado, John W. Galbraith, and David F. Hendry. 1993. *Co-*

integration, Error Correction, and the Econometric Analysis of Non-stationary Data. Oxford University Press.

Bewley, Ronald A. 1979. The direct estimation of the equilibrium response in a linear dynamic model. *Economics Letters* 3(4): 357–361.

Box-Steffensmeier, Janet M., John R. Freeman, Matthew P. Hitt, and Jon CW Pevehouse. 2014. *Time Series Analysis for the Social Sciences*. Cambridge University Press.

Burke, Simon, John Hunter, and Alessandra Canepa. 2017. *Modelling Non-stationary Economic Time Series: A Multivariate Approach*. Palgrave.

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

Durr, Robert H. 1993. What Moves Policy Sentiment? *American Political Science Review* 87(1): 158–170.

Edwards, George C. 2009. *The Strategic President: Persuasion and Opportunity in Presidential Leadership*. Princeton University Press.

Ferguson, Grant P., Paul M. Kellstedt, and Suzanna Linn. 2013. How Does the Economy Shape Policy Preferences? *Electoral Studies* 32(3): 544–550.

Gelfand, Alan, Susan E. Hills, Amy Racine-Poon, and Adrian F. M. Smith. 1990. Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association* 85(412): 972–985.

Grant, Taylor and Matthew J. Lebo. 2016. Error Correction Methods with Political Time Series. *Political Analysis* 24(1): 3–30.

Hill, Jonathan and Liang Peng. 2014. Unified Interval Estimation for Random Coefficient Autoregressive Models. *Journal of Time Series Analysis* 35(3): 282–297.

Hill, Jonathan, Deyuan Li, and Liang Peng. 2016. Uniform Interval Estimation for an AR(1) Process with AR Errors. *Statistica Sinica* 26(1): 119–136.

Hopkins, Vincent, Ali Kagalwala, Andrew Q. Philips, Mark Pickup, and Guy D. Whitten. 2023. How Do We Know What We Know? Learning from Monte Carlo Simulations. *Journal of Politics*. 86(1): 36–53.

Keele, Luke, Suzanna Linn, and Clayton McLaughlin Webb. 2016. Treating Time with All Due Seriousness. *Political Analysis* 24(1): 31–41.

Lebo, Matthew J., and Andrew J. O'Green. 2011. The President's Role in the Partisan Congressional Arena. *Journal of Politics* 73(3): 718–734.

McNeish, Daniel. 2016. *On Using Bayesian Methods to Address Small Sample Problems. Structural Equation Modeling: A Multidisciplinary Journal* 23(5): 750–773.

Miočević, Milica, Roy Levy, Rens van de Schoot. 2020. Introduction to Bayesian Statistics. In *Small Sample Size Solutions*, eds van de Schoot, Rens, and Milica Miočević. Routledge. 1–10.

Murr, Andreas, Richard Traunmüller, and Jeff Gill. 2023. Computing Quantities of Interest and Their Uncertainty using Bayesian Simulation. *Political Science Research and Methods* 11(3): 623–632.

Pesaran, M. Hashem, Yongcheol Shin, and Richard J. Smith. 2001. Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics* 16(3): 289–326.

Peterson, David AM, Joanne M. Miller, Kyle L. Saunders, and Scott D. McClurg. 2022. Macrointerest. *British Journal of Political Science* 52(1): 200–220.

Philips, Andrew Q. 2018. Have Your Cake and Eat It Too? Cointegration and Dynamic

Inference from Autoregressive Distributed Lag Models. *American Journal of Political Science* 62(1): 230–244.

Pickup, Mark. 2022. Equation Balance in Time Series Analysis: Lessons Learned and Lessons Needed. *Political Science and Research Methods* 10(4): 890–900.

Pickup, Mark and Vincent Hopkins. 2022. Transformed-likelihood Estimators for Dynamic Panel Models with a Very Small T. *Political Science Research and Methods* 10(2): 333-352.

Pickup, Mark and Paul M. Kellstedt. 2022. Balance as a Pre-Estimation Test for Time Series Analysis. *Political Analysis* 31(2): 295–304.

Stimson, James A. 2018. *Public Opinion in America: Moods, Cycles, and Swings.* 2nd. edition. Boulder, CO: Westview Press.

Webb, Clayton, Suzanna Linn, and Matthew Lebo. 2019. A Bounds Approach to Inference Using the Long Run Multiplier. *Political Analysis* 24(1): 281–301.

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

Wlezien, Christopher. 1995. The Public as Thermostat: Dynamics of Preferences for Spending. *American Journal of Political Science* (1995): 981-1000.