Long-Run Confidence: Estimating Uncertainty when using Long-Run Multipliers

Mark David Nieman
mark.nieman@utoronto.ca
Sidney Smith Hall, Room 6026B
100 St. George Street
Toronto, ON M5S 3G3

David A.M. Peterson
daveamp@iastate.edu
541 Ross Hall
527 Farm House Ln
Ames, IA 50011

Abstract

Researchers are often interested in the long-run relationship between variables where the dependent variable has dynamic properties. Though determining the long-run multiplier (LRM) for an independent variable is straightforward, correctly estimating the significance of the LRM is often difficult, especially when time series are short and tests for series' stationarity are uncertain. We propose a Bayesian framework for estimating the LRM by using a bounded prior on the lagged dependent variable to constrain estimates for dynamic processes 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 provides direct estimates of the LRM and its uncertainty, even for short time series. We highlight the advantages of this approach via Monte Carlo experiments and replicate several studies to show that our method clarifies long-run relationships that were inconclusive using existing techniques.

Time series researchers face difficult choices. Determining the best, let alone correct, specification is challenging. In most cases, the first step is to test the stationarity of our series and make our modeling choices based on these diagnostics. Philips (2018), for instance, provides a useful flowchart of choices that one should make based on the results of the 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 researcher, then, has to do what they think 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 believe 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) summarize the issues that time series analysts face and introduce the associated bounds testing procedure of Pesaran, Shin, and Smith (2001) (PSS) to political science. PSS's approach recognizes the uncertainty we often have about the stationarity of our independent variables. Unfortunately, these early discussions still treat the diagnostics of 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) which summarizes the relationship between each independent variable and the dependent variable. They show that the significance of the LRM is a test of the presence of the long-run relationship (LRR) between the variables, regardless of stationarity. Importantly, they also provide the bounds of the t-test of the LRM that researchers can use to infer the statistical 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 its uncertainty, and then the ratio of the 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, both of which are asymptotically equivalent to the direct estimation of the standard error of the LRM. WLL show that the distribution of the ratio of the LRM and its standard error is, however, 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 can lead to frustration for applied researchers. The bounds method that WLL use has a range of values where the hypothesis of an LRR between X and y is rejected, a range where it is not rejected, and a range of values that is indeterminate.[1] Their advice is to treat results that fall in this indeterminate range as failing to reject the null hypothesis of no relationship and to 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 can be somewhat mitigated. In this manuscript, we develop a simple Bayesian estimator of the LRM that does not have this indeterminacy. We start by using a bounded, uniform 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

[1]We follow WLL and denote X as a set of multiple regressors and x to indicate a single regressor—our focus are models with a single dependent variable.

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.

Bayesian estimation with a semi-informed prior to estimate uncertainty for the LRM offers a number of benefits. The semi-informed prior accommodates series of X and y with unclear dynamic properties by limiting the range of the coefficient on y_{t-1} to its theoretical bounds but by giving equal density to the values between these bounds, so as not to bias point estimates. Moreover, the use of a sampling-based method, like MCMCs, allows for *direct* estimation of the variance of the LRM, without requiring large sample sizes. Our proposed method leads to more accurate and reliable estimates of uncertainty than alternatives that rely on asymptotic assumptions that may not hold.

In the next section, we revisit the results presented by WLL, demonstrating the importance of the significance tests of the LRM. Next, we describe our approach. We demonstrate its advantages using two Monte Carlo experiments and two empirical applications—one included in WLL (2019) and a recent publication—and conclude.

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. The key to these models is the existence of an LRR between X and y, which implies a long-run equilibrium between the two. The presence of an equilibrium means that the variables tend to not change over time, but the practical implication is that when they do deviate, they tend to return to the equilibrium state (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 equilibrium is the mean, i.e., the series will revert to the mean post-deviation. The type of the equilibrium also depends on the relationship 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 an *unconditional stationary equilibrium*. In contrast, a variable that is non-stationary, by definition, does not have an equilibrium level 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 they 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 used for our hypotheses and the critical values of those tests depend on these diagnostics. Getting the diagnostics 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 should an applied researcher move forward? If we knew the type of equilibrium possible for our variables, then we would know which model to use. The flowchart in Philips (2018) provides clear guidance on this. If X and y are all stationary, run an autoregressive distributed lag (ADL) model.[2] If the autoregressive distributed lag-bounds test suggests cointegration, estimate a GECM. If there is not enough evidence to conclude that there is

[2]There is a fair amount of inconsistency in the use of acronyms for these models. PSS use "ARDL" to refer to a model like a GECM where a unit root y is differenced in the equation. The political science literature has tended to use ADL to refer to the type of equation we are referencing here. We use ADL only to refer to equations where the y variable is included in levels.

cointegration, difference the variables and then run an ADL. This advice is straightforward and helpful, and the bounds approach created by PSS is an excellent step forward. Jordan and Philips (2018a,b, 2020) have also made packages in both R and Stata available to implement these methods.

The problem with this approach is, as WLL note, that it starts with the assumption that one can definitively diagnose whether the dependent variable is stationary. This is often much harder than it sounds. Unit root tests have low power, particularly with the short time series that are common in political science (Lebo and Kraft, 2017). 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. Therefore, they must hope that the results are robust enough to these numerous specification choices, that their conclusion is invariant to the chosen approach.

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

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

where the LRM is represented as λ and 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; when both variables are stationary, the equation is balanced (Pickup and Kellstedt, 2022). 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 as an ADL model:

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

or via the GECM:

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

where $\alpha_1^* = (1 - \alpha_1)$, $\beta_0^* = \beta_0$, and $\beta_1^* = \beta_0 + \beta_1$. This model choice is a matter of taste as they are mathematically equivalent (Marriott and Newbold 1998: 327–328; De Boef and Keele, 2008: 189–190). Regardless of the specific model used, of particular interest is the LRM, which is calculated as $\frac{\beta_0+\beta_1}{1-\alpha_1}$ for the ADL and $-\frac{\beta_1^*}{\alpha_1^*}$ in the GECM.

While recovering a point estimate of the long-run relation by simply inputting the estimated coefficient into the appropriate formula is relatively straightforward, calculating uncertainty is more complicated. As De Boef and Keele (2008) note, neither the ADL 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 to *approximate* the variance of this ratio and estimate uncertainty in the LRR. The first is to calculate 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 a constant, x_t, x_{t-1}, and y_{t-1} are used as instruments (De Boef and Keele, 2008: 192). 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. Most statistics packages have made the estimation of the standard error straightforward.[3] Still, while these estimates of the standard error are *asymptotically* accurate, they may not be as appropriate in the small sample sizes typical in applied time series work.

WLL demonstrate convincingly that the significance of the LRM is a clear indication of an LRR between X_t and y_t. As they note "*a nondegenerate, or valid, equilibrium relationship between y_t and x_t requires the LRM to be nonzero*" (Webb, Linn, and Lebo, 2019: 286-287, emphasis in original). They demonstrate that this holds regardless of the stationarity of y_t, which helps resolve much of the uncertainty in pre-analysis specification tests of the 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 LRR with y_t. Other approaches, like the Granger-Engle two-step method or the autoregressive distributed lag-bounds approach, only indicate that a cointegrating relationship exists between the dependent variable and *at least one* of the independent variables, but not which one.

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 depends on the sample size and the degree of autoregression in y_t and x_t, in general, they find that the critical values do not 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 overconfident. More importantly, they develop the bounds for the hypothesis tests of the LRR.[4]

[3] E.g., the `nlcom` command in Stata and the `deltamethod()` command in the `CAR` package in R.

[4] Webb, Linn, and Lebo (2020) provide an expanded set of critical values to denote upper and lower

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 frustrating 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.[5] Some of the applied work that relies on WLL also treats results in this indeterminate region as if they fail to reject the null (e.g., Wolak and Peterson, 2020). Yet, an LRR may exist even if estimates fail to reach significance owing to several reasons, including low power and a lack of precision (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 series. A lack of precision may result from measurement error in one of the variables or from a short series that is truncated before returning to equilibrium. In either case, when the number of observations is small, the estimate of α_1, the coefficient on y_{t-1}, can be imprecise. In cross-sectional models, these issues simply result in inflated standard errors. With dynamic processes, however, the inflation of standard errors, amplified through y_{t-1}, also affects the uncertainty of the LRM. 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.

Given these issues, common to social science data, researchers face numerous practical difficulties. The upper bound criterion of the bounds test may be too stringent for identifying LRRs in short time series or when a researcher's data simply do not play nice and definitively meet (or violate) stationarity assumptions, based on the usual battery of checks.

bounds, by both number of observations [25, 50, 75, 150, 500, 1000] and α-level [.01, .05, .10].

[5] See Webb, Linn, and Lebo (2019: 293, 299) and Webb, Linn, and Lebo (2020: 286).

A Bayesian Approach

We propose a Bayesian solution to the problem.[6] Specifically, we formalize the range of plausible values of the dynamic relationship between independent and dependent variables with the use of a semi-informed prior. This formalization—whether paired with other informed or non-informative priors on other parameters—allows for constructing measures of uncertainty for the LRM, whether or not one adopts a fully Bayesian theoretic and inferential perspective. Our approach is general, but especially advantageous for short series.

We start by providing an overview of Bayesian inference, with a focus on application to time series analysis. Bayesian statistics is based on Bayes' Law: $P(\theta|data) = \frac{P(\theta)P(data|\theta)}{P(data)}$, where $P(\theta|data)$ is the posterior probability, $P(\theta)$ is the prior probability, $P(data|\theta)$ is the probability density/mass function (e.g., a likelihood), and $P(data)$ are the observed data. Within this framework, the focus is on probabilistic statements using distributions on all facets: prior distribution, posterior distribution, and even the data itself. The data in the denominator, for example, have been observed and thus have a probability equal to one. Bayesian analysis, in other words, treats observed data as given and asks what distribution is the most likely to have produced it.[7] This view lends itself to a probabilistic treatment of parameter estimates and favors model comparison. In this respect, Bayesian analysis closely resembles maximum likelihood (ML), as random variables are treated as conditioned

[6] Within political science, Bayesian tools have become a standard part of a researcher's toolkit, being applied to a number of tasks, e.g., multilevel models (King and Gelman, 1991; Western, 1998), measurement models (Jackman, 1994, 2001; Quinn, 2004; Schnakenberg and Fariss, 2014), matching administrative data (Enamorado, Fifield, and Imai, 2019), text analysis and topics modeling (Grimmer, 2010; Roberts et al, 2014; Eshima, Imai, and Sasaki, 2024), modeling high-order network dependencies (Hoff and Ward, 2007; Minhas, Hoff, and Ward, 2019), forecasting political events (Ward, Greenhill, and Bakke, 2010; Mueller and Rauh, 2018), model averaging (Montgomery and Nyhan, 2010), and missing data (Honaker and King, 2010; Hollenbach et al, 2021). Bayesian methods are also frequently used within time series analysis, particularly for non-stationary data—e.g., change-point (Spirling, 2007; Park, 2010, 2011, 2012) and vector autoregressions models (Brandt and Freeman, 2006, 2009; Brandt and Sandler, 2012)—and autoregressive processes with complex time-series cross-section error structures (Pang, 2010, 2014). Substantive fields applying Bayesian time series methods include election forecasting (Linzer, 2013) and voting behavior (Peterson, 2009), democratic backsliding (Knutsen et al, 2024), human security (Brandt and Sandler, 2010; Santifort, Sandler, and Brandt, 2013; Fariss, 2014), international conflict (Brandt, Colaresi, and Freeman, 2008; Park, 2010; Nieman, 2016), and major power competition (McGinnis and Williams, 1989; Thies and Nieman, 2017), among others. See Gill (2012) and Park and Shin (2020) for reviews.

[7] See Gelman et al (2004), Gill (2009), and Chan et al (2020) for an introduction to Bayesian analysis.

by the observed data to offer the 'most likely' outcomes (Gill, 2009: 39–44); in fact, Bayesian and ML models are asymptotically equivalent, regardless of the specified prior distribution (Gelman et al, 2004: 111, 247; Gill, 2009: 4, 60–61, 62–65).[8] Conversely, the frequentist approach treats the observed data as a random variable arising from a fixed parameter. That is, a distribution is assumed (the null hypothesis) and the question is whether the data differ from it. This view lends itself to null hypothesis testing, with estimates of parameters used to either reject or fail to reject the null.

This difference in how the parameters and data are treated has two key implications when applied to time series analysis. The first relates to the classification problem when determining the characteristics of the observed data, e.g., stationary, non-stationary, cointegration, etc. As described above, the uncertainty associated with the diagnostic pre-tests makes classifying y, X, and the equilibrium LRR a non-trivial task. The source of much of this difficulty arises from treating the data as a random variable, where the observed data are manifestations of an unobserved underlying data generation process. Hence, the data themselves are uncertain, as are their properties.

From a Bayesian perspective, however, this is less problematic, as the data are known (Chan et al, 2020: 343–344). Therefore, pre-testing the data to reject or fail to reject a null hypothesis about their data generating process does not make sense: the data are known with a probability of 1 and it is the model parameters that are unknown. Rather than a pre-test for the presence or absence of a unit root, one could instead directly compare two competing model specifications, one with a unit root and one autoregressive, and make a probabilistic assessment of their fit to the data using Bayes Factor—a ratio of the models' marginal likelihood—as evidence (Marriott and Newbold, 1998). Whereas pre-testing lacks clarity in the face of uncertain or conflicting diagnostic results, model comparison provides greater direction along with the degree of confidence in the selected choice.[9]

[8] A key difference, however, is that ML estimation maximizes only the likelihood, while Bayesian estimation via MCMCs approximates the entire posterior distribution (Gill, 2009: 19, Miočević, Levy, and van de Schoot, 2020: 7); we take advantage of this feature when estimating uncertainty in the LRM.

[9] Marriott and Newbold (1998) show that Bayes Factor can distinguish unit root and autoregressive

A second benefit of the Bayesian inferential model is that it works well for time series with small T. Whereas maximum likelihood and frequentist approaches rely on the central limit theorem, Bayesian analysis does not. Instead, Bayesian analysis typically applies sampling-based methods, such as MCMCs, where the quality of inference relies on the *number of samples* taken rather than the *size of the sample* (McNeish, 2016). This distinction is most pronounced in the estimates of uncertainty when the number of observations is small. This feature is useful for applied time series analysis, as many social science series are short, which may cause diagnostic problems.[10]

The main advantage of a Bayesian framework, for our purposes, lies in leveraging a semi-informed prior to help estimate uncertainty in the LRM. The prior can be specified so as to restrict the coefficient of the lagged DV within the theoretical range of the estimator: strictly between -2 and 0 in the GECM and -1 and 1 in the ADL.[11] We specify a diffuse, uniform prior that places equal probability on all values between these bounds. Alternatively, in settings where one can incorporate existing knowledge, a more informed prior may be more practical, e.g., one that places more weight on values closer to 0 or either bound.[12]

One way to think about the semi-informed prior is that it is simply the formalization of a 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, they assume that the root of the characteristic equation of the time series is less than one. When the series is integrated,

processes, even for short series where α_1 takes on high values. Moreover, one could use Bayes Factor or BIC to apply a general-to-specific modeling approach. For instance, Bayes Factor is commonly used to differentiate models in applications using change-point models.

[10] As a completely uninformative prior may increase bias and reduce coverage in small samples, weakly informed priors that capture known theoretical characteristics—such as the one we propose—are recommended. (McNeish, 2016; Veen and Egberts, 2020).

[11] See Keele, Linn, and Webb (2016: Table 1) for a summary of error correction rates and long-run equilibria.

[12] If one is estimating an ADL, for example, and has theoretical reasons to expect the coefficient on the lagged DV to approach 1, then a sharp prior where the pmf is massed near 1, such as $\mathcal{B}(5, .5)$, could be used (see Marriott and Newbold, 1998: 328–333). If, instead, a researcher believes that the coefficient on the lagged DV is positive, but has no other information, then a $\alpha_1 \sim \mathcal{B}(1, 1)$ prior would apply a uniform distribution between 0 and 1.

the root of the characteristic equation is exactly one. The use of the semi-informed 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.

Using a bounded prior also keeps the estimates of uncertainty for the LRM firmly within their theoretical limits. If the confidence interval for α_1 nears or exceeds 1, then the denominator of the LRM can take very small, or even negative, values. In that case, the estimation of the variance of the LRM will be "mildly explosive."[13] The prior thus keeps estimates of uncertainty within the same theoretical bounds as the point estimate, providing more substantively plausible and theoretically-informed results.

The actual estimation of the Bayesian model is carried out via MCMCs. This allows us to calculate the distribution of the posterior for all of the coefficients directly, including the LRM. Once the model has converged, each simulation of the MCMC draws all of the parameters in the model from their joint probability distribution. The ratio of the parameters in the draw is, then, also a draw from the posterior of that ratio (Gelfand et al, 1990; Murr, Traunmüller, and Gill, 2023). Table 1 illustrates this for an ADL model, specified as in Equation 2. For the purposes of the illustration, assume that $\alpha_0 = 0$, $\alpha_1 = 0.5$, $\beta_0 = 0.25$, and $\beta_1 = 0.5$, resulting in the LRM, $\frac{0.5+0.25}{1-0.5} = 1.5$, and that the MCMC has 5,000 iterations. The table shows that for each individual simulation, we recover parameter estimates of the specified model. For each individual iteration, then, we can construct and calculate the LRM based on these coefficients.

As is standard with Bayesian estimation, the posterior distribution can be summarized to construct point estimates (e.g., median) and uncertainty (e.g., 95% credible region), as well as indicate the percent of individual draws above or below zero, for each model parameter. These summaries can also be constructed for the LRM. As such, we do not need to rely on an asymptotic equivalent to the confidence interval of the LRM, such as the Bewley transformation or delta method. In cases where a time series is relatively short, this direct estimate

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

Table 1: Hypothetical parameter estimates for Markov chain Monte Carlo.

Iteration	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\widehat{\text{LRM}}$
1	0.02	0.64	0.43	0.36	2.19
2	-0.05	0.38	0.40	0.15	0.89
3	-0.13	0.57	0.71	0.26	2.26
⋮	⋮	⋮	⋮	⋮	⋮
5000	-0.15	0.52	0.62	0.11	1.52

Note: Long-run multiplier (LRM) calculated from an autoregressive distributed lag model.

of the uncertainty of the LRM should be more accurate and reliable than approximations that rely on asymptotic properties that may not hold.

To demonstrate the advantages of our approach for estimating uncertainty in the LRM, we conduct two Monte Carlo experiments. The first experiment compares coverage rate estimates for the LRM—using both the bounds approach from a Bewley transformation of the ECM, and a Bayesian ECM with a semi-informed prior—under varying levels of univariate autocorrelation of x and y and for different time series lengths, when there is no LRR. These conditions allow for assessing how well each approach separates a spurious long-run bivariate relationship from univariate dynamics.

The second experiment illustrates the application of our approach to a more realistic (but still controlled) setting where the researcher is interested in 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, in the presence of a moderate LRR between x and y and a moderate autocorrelation between x and its lag. This specification allows us to evaluate how well each approach detects a true LRR under plausible conditions. Following these experiments, we compare the results and provide a general discussion.

Monte Carlo Experiment #1

In this experiment, we generate our data following the dynamic simulation process used by WLL. 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 $T = 1000$. 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 $\rho_y = 1$ and $\rho_x = 1$. In the true data generating process there is no LRR 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 transformation from Eq 5. We also estimate a Bayesian ECM. The Bayesian ECM is specified as in Eq 3, with diffuse priors of $\mathcal{N}(0, 20)$ for the constant and the coefficients associated with Δx_t and x_{t-1}, a prior of $\mathcal{U}(-2, 0)$ on the coefficient for lagged y, and a 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 specifications for two lengths of T, such that $T \in \{25, 75\}$.[14] The varying lengths for T allow us to look at both how well the estimates fare for short and moderately short time series that are common to social science data.

Table 2 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).[15] 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. The point estimate gives the expected LRM, while the coverage rates offer insights

[14]Both lengths are from the same time series and the smaller length is a subset of the longer one.

[15]To clarify, within each individual simulation of this Monte Carlo experiment, Bayesian ECM estimates are summaries of a posterior distribution of 5,000 MCMCs.

Table 2: Long-run multiplier 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	-0.01	-0.01	0.01	-0.01
Coverage	0.96	0.96	0.89	0.90
Bewley transformation				
Mean	0.01	-0.01	-0.01	0.09
Lower bound coverage	0.76	0.75	0.78	0.65
Upper bound coverage	0.99	0.99	0.99	0.95
Indeterminate range	0.24	0.25	0.22	0.30
	$T = 75$			
	$\rho_y = 0, \rho_x = 0$	$\rho_y = 0, \rho_x = 1$	$\rho_y = 1, \rho_x = 0$	$\rho_y = 1, \rho_x = 1$
Bayesian ECM				
Median	0.01	-0.01	-0.01	-0.01
Coverage	0.95	0.95	0.88	0.88
Bewley transformation				
Mean	-0.01	-0.01	-0.01	-0.25
Lower bound coverage	0.70	0.70	0.76	0.58
Upper bound coverage	0.99	0.99	0.99	0.95
Indeterminate range	0.30	0.30	0.24	0.37

Note: Coverage for the Bayesian error correction model (ECM) based on 95% credible intervals from the posterior distribution. Coverage for the Bewley transformation calculated using t-statistics of 1.25 (lower bound) and 3.79 (upper bound) when $T = 25$, and t-statistics of 1.06 (lower bound) and 3.73 (upper bound) when $T = 75$.

into how well the different estimation strategies perform at recovering accurate estimates under varying conditions (Hopkins et al, 2023). Lower coverage rates would suggest that, even if an estimator is unbiased, on average, its results are less reliable in any particular application. Coverage rates for the Bayesian ECM report how frequently the true value is within the estimated 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).[16] We also report the percentage of indeterminate cases arising from incongruent outcomes between using lower and upper bounds.

The Monte Carlo experiment provides four key insights. Given our interest in estimates

[16] 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 3.73.

of uncertainty, we focus on coverage rates.[17] 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 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, with the aim of the WLL's bounds approach). The lower bound, however, performs much worse, with a high coverage rate of 78% and a low of 58%. The indeterminate range, where researchers cannot either confidently reject the null hypothesis or fail to reject it, is never less than 21% and reaches 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 size properties (McNeish, 2016; van de Schoot and Miočević, 2020). In contrast, the Bewley coverage for the upper bound remains consistent, whereas that of the lower bound actually decreases as T increases. This stems from a smaller t-statistic being applied when using the bounds approach as the number of observations increases. As a result, the proportion of cases within the indeterminate range between the bounds increases with sample size.

Third, the percentage of cases in the indeterminate range for the bounds approach estimates is greatest when $\rho_x = 1$. This result is substantively meaningful, as many independent variables common to panel data either do not change, or do not change very much, over time. For instance, country-level institutional features, e.g., regime type, are usually stable for long periods of time. Even covariates that change, e.g., GDP, are often primarily a function of their own prior value. Each of these is a case 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

[17]Each approach returns similar point estimates of the LRM, which is unsurprising given they both are estimated using OLS.

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 toward 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 the two approaches 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[18] while the indeterminate range for the bounds approach remains large.

Monte Carlo Experiment #2

For this Monte Carlo experiment, we generate a simple dynamic model with a moderate LRR between x and y that mimics common features of real-world 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 detects the LRR among variables of interest.

We recover parameter estimates using a Bayesian ECM and a traditional ECM, specified as in Eq 3. The Bayesian ECM estimates the LRM directly from the posterior distribu-

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

tion, and for the traditional ECM, we estimate the LRM using the Bewley transformation, specified as in Eq 5. 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 be non-negative), 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.

Table 3 summarizes the results based on 20,000 simulations. The first column reports the true value of the ECM estimates based on the data generating process described above;[19] 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 parameters 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 range of indeterminate values where the lower and upper bounds give conflicting inferences.

As both the Bayesian and traditional ECM return similar parameter estimates—including for the LRM—we again focus on coverage rates. The entirety of the Bayesian estimator's 95 percent credible interval range is greater than zero in 68% of the simulations (the rest include 0).[20] In contrast, the 95 percent confidence intervals estimated using the bounds approach offer a more muddled conclusion. 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 simulations, the more challenging upper bound criteria does so only 29% of the time. The result is that, in almost two-thirds of the simulations (65%), the bounds approach offers conflicting guidance on whether an LRR exists between x and y; this is even though a true

[19]Recall that the ADL and ECM are mathematically equivalent.

[20]For enhanced comparison, we report coverage rates for the entirety of the credible interval. One could alsoreport the percent of individual MCMC draws above/below zero so as to better convey the confidence regarding a LRR between X and y. We demonstrate this in the next section.

Table 3: Error correction model parameter estimates with a moderate long-run relationship, univariate autocorrelations, and $T = 25$.

	True value	Bayesian ECM	ECM w/ Bewley trans. & bounds
y_{t-1}	-0.50	-0.58	-0.59
		[-0.88, -0.25]	(0.17)
X_{t-1}	0.75	0.80	0.81
		[0.22, 1.37]	(0.27)
ΔX	0.50	0.49	0.50
		[0.01, 0.98]	(0.22)
Constant	0.00	-0.01	-0.01
		[-0.51, 0.51]	(0.24)
Long-run multiplier	1.50	1.45	1.45
		[0.17, 3.47]	(0.54)
Bayesian coverage rate		0.68	
Lower bound coverage			0.94
Upper bound Coverage			0.29
Percent in indeterminate range			0.65

Note: Coverage for the Bayesian error correction model (ECM) based on 95% credible interval from the posterior distribution. Coverage for the Bewley transformation calculated using a t-statistic of 1.25 (lower bound) and 3.79 (upper bound).

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 $I(0)$ nor $I(1)$, but in-between.

There are several takeaways from the two experiments. One is that the bounds approach—especially its stricter criterion of using only the upper bound—reduces the risk of 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 hand, is much better able to correctly recover moderate LRRs, 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. 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 are 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 summarizing the data and manageable with even a slightly-informed prior—one gains far more precision in making theoretical inferences.

Applications

We replicate three existing time series papers using our approach. We begin with the recent paper on aggregate levels of interest in politics in the U.S. by Peterson et al (2022). The specifics of the LRM in this application provide good illustrations of the above points. Following WLL (2019), we also re-estimate applications from Lebo and O'Green (2011) on presidential success in Congress and Ferguson, Kellstedt, and Linn (2013) on the effect of policy on public mood. The first two applications are reported in this section, with the third in the Supplemental Materials. These applications demonstrate that the Bayesian approach can resolve some of the indeterminate results in existing work.

Replicating Peterson et al., 2022

Peterson et al (2022) argue that trust in government and interest in politics trade-off in the electorate. When the electorate trusts the government more, the incentive to pay attention to politics lessens. Voters choose 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. The authors expect that when the electorate trusts the government to look out for the public, the incentive to actively monitor the government lessens. Some of the electorate choose to spend that time on things they find more enjoyable than politics. In contrast, when 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: while higher levels of trust and higher levels of interest are believed to be markers of a stronger democracy, it is difficult to have high levels of both.

The authors rely on existing macro-level measures of trust in government and use the technique developed by Stimson (2018) to construct an 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 indicates that they are cointegrated. Based on this, they estimate 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 found no evidence for either relationship. Peterson et al (2022) use the GECM setup for their model of macrointerest. A simplified version of their specification is:

$$\Delta y_t = \alpha_0 + \alpha_1^* y_{t-1} + \beta_1^* \Delta x_1 t + \beta_2^* x_{1t-1} + \beta_3^* \Delta x_2 t + \beta_4^* x_{2t-1} + \beta_5^* \Delta x_3 t + \beta_6^* x_{3t-1} + \epsilon_t \quad (6)$$

where y_t is macrointerest at time t, x_1 is trust in government, x_2 is consumer sentiment, and x_3 is presidential approval.[21]

The conclusion Peterson et al (2022) reach about the effect of trust on interest, however, is dependent on their diagnostic analyses 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 uncertain about the assumption that both trust and interest have unit roots.

We re-analyze this study using our approach. We also take advantage of this application to provide a step-by-step guide for implementing our approach in R using the `brms` package

[21]The model also includes several indicators for specific political events and a measure of presidential campaigns. None of these controls are expected to have a LRR with macrointerest and we omit them from the equation for brevity, but do include them in the model to fully replicate the original work.

(Bürkner, 2017).[22] The `brms` package excels at more complicated Bayesian models, but the specification of our model is quite straightforward. The `brm` command is specified as most other linear models in R, with the designation of the dependent and independent variables.[23] The `brm` command requires the specification of the family of the model (in this case, Gaussian), the number of chains for the MCMC (we use 4), the number of iterations (300,000), the burn-in (250,000), and the thinning rate of the MCMC sampler (5). These are all standard features of an MCMC model and the simplicity of the estimation makes these choices moderately trivial.

The only part of the model specification that requires any real choice is that of the priors, but even this is straightforward. The `brm` command includes default priors for all parameters in the model: these will be used for any parameter with no user-specified prior. For this type of model, the priors for the β parameters with the independent variables are flat. The prior for α_0 is specified as a Student-$t_{(3,0,2.5)}$ and the prior on σ is specified as a Half Student-$t_{(3,0,2.5)}$. The user can change these defaults, but our approach does not require any changes. The only user-specified prior in the model is on the α_1 parameter, where we specify a $\mathcal{U}(-2,0)$ prior.[24]

This creates a brms-class object in R, from which the summary command will return, among other things, the parameter estimates, their estimated error, and the 95 percent credible interval. Users should be sure to check the standard MCMC diagnostics to ensure that the chains have converged on the posterior, though these models are generally not computationally-complex enough to cause such problems. The object will also contain the draws from the posterior. The code `as_draws_df` will let the user define a new dataframe made up of these draws. The user can then construct the LRM of any independent variable of interest. The hypothesis tests are a simple matter of summarizing this calculated LRM.

[22] Analogous steps may be followed in Stata using the `bayesmh` command.

[23] `brms` does not have a direct way to lag and difference variables, so data processing occurs prior to model specification.

[24] This is done with a single line of code: `prior = prior(uniform(-2, 0), coef = ldv)`, where ldv is the lagged DV.

In our applications, we calculate the 95 percent credible interval and the proportion of draws of the LRM that are either less than or greater than zero (depending on the direction of the hypothesis). The reporting of the proportion of draws, in particular, conveys a more precise and accurate level of confidence in whether a variable has a non-zero effect than is garnered from frequentist hypothesis testing.

The results, reported in Table 4, have the same pattern of significance as the original article. The coefficients are medians from the posterior of the MCMC, the numbers in the parentheses underneath are their 95 percent credible intervals. There is a negative relationship between trust and interest in both the LRR and the short-term effect of changes in trust in government. Central to our application, the 95 percent credible interval of the LRM excludes zero, suggesting that there is an LRR 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. Lastly, and consistent with Peterson et al (2022), we find no evidence that either the index of consumer sentiment or presidential approval is linked to macrointerest.

We can also plot the distribution of parameters to illustrate the "mildly explosive" potential of the variance estimates if the error correction rate is too slow. Figure 1 presents the density plot of the draws from the posterior for the coefficient on the lagged interest measure (the error correction rate) and the lagged trust measure. The posterior of the lagged interest measure is a normal distribution and the density plot looks like draws from a typical normal distribution, with the peak centered at the mean/median. The one thing to note is that the posterior for interest does have a small number of draws that are very close to zero. The posterior of the lagged trust measure also looks as expected and simply visualizes the median and credible interval presented in Table 4. Some of the draws from the posterior are near zero, but that will not cause the mildly explosive issue in the standard error of the LRM.

Next, Figure 2 presents the distribution of the LRM for the effect of trust on interest, the key hypothesis in the paper. Note the range on the x-axis. The posterior distribution has a

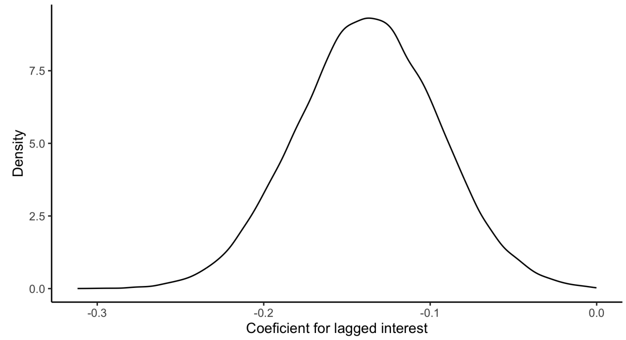
Table 4: A model of macrointerest, 1973–2014.

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		

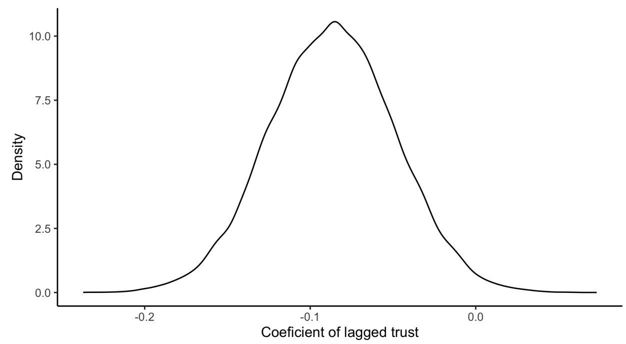
Note: Bayesian error correction model replicating Peterson et al. (2022). Results present the median and 95 percent credible region of the parameter estimates from the posterior distribution. LRM = long-run multiplier.

handful of extreme values (both positive and negative). The vast majority of the mass of the distribution is around the median value of -0.62, but some are several standard deviations away from the median of the distribution. This is because a small number of the draws from

Figure 1: Estimated posterior of the lagged values of interest and macrointerest.
a) Posterior distribution of the coefficient for lagged interest.



b) Posterior distribution of the coefficient for lagged trust.

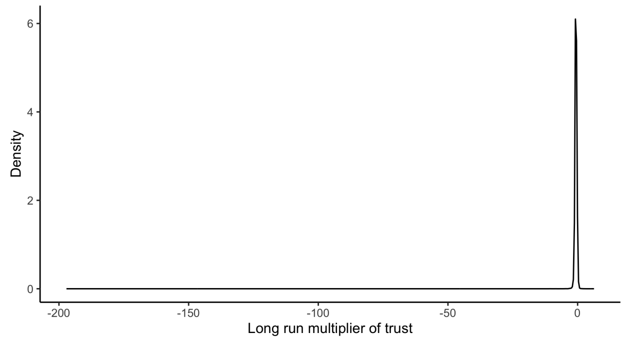


the posterior of the α_1 parameter are very close to zero.

This illustrates the claim by Hill and Peng (2014, 293) and Hill, Li, and Peng (2016, 126) that if the variance of the α_1^* parameter has probability mass at 0, the variance estimate of the LRM will be mildly explosive. That is, because α_1 is in the denominator of the LRM, if its value in the draw from the posterior is too close to zero, the absolute value of the LRM for that draw will be extraordinarily high, regardless of the coefficient on the independent variable in that same draw. In this application, there is a draw from the posterior of the LRM for trust that is -72.80, almost one hundred times the median estimate and more than seventy times the lower value of the credible interval. This is because the estimate of α_1 in that draw is -0.00008. Even though the estimate of the coefficient for the lag of trust (-0.06) is below the median value of the posterior, the calculation of the LRM for that draw is an extreme value.

In this application, only a handful of the draws from the posterior of α_1^* are close enough to

Figure 2: Estimated posterior of the coefficient for the long-run multiplier for trust.



zero for this issue to occur. Of course, as shown with the 95 percent credible interval reported in Table 4, these outliers are not common enough to substantially affect the interval's width, but do induce a slight asymmetry and left skew. This result is intuitive and correctly reflects that, in this specific application, the error correction rate is slow and any shocks cause the series to return to equilibrium very slowly. If, however, the estimate of α_1 was closer to zero, meaning that the error correction rate in the process was even slower, then there would be an even longer tail and a more skewed credible interval.

This also points to an advantage of a theoretically-bounded prior: plotting substantive effects from a model with unbounded estimates of α_1, whether estimated using a frequentist or Bayesian model, could result in standard errors above their theoretical limits, longer tails, and substantive effects that include values of infinity among their draws—making their uncertainty intervals, or potentially even their point estimates, nonsensical. For example, when plotting substantive effects of the LRM estimated from a frequentist model using conventional simulation methods—i.e., drawn from a multivariate normal distribution—some draws would likely include values near or below zero in the denominator. Likewise, a Bayesian model without a semi-informed prior would face similar problems, just at the estimation stage when constructing the LRM from the MCMCs.

Replicating Lebo and O'Green, 2011

Next, we replicate the work of Lebo and O'Green (2011), which 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 research question is whether presidential approval gives the president leverage 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 influence, 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.

The application contains three independent variables explaining presidential success. The main variable is presidential approval. The other two variables are the percentage of House seats held by the president's party and 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 a straightforward ECM model.

WLL find evidence for the effect of the share of the House held by the president's party and no evidence of the effect of presidential approval. The t-statistic capturing the relationship between CPG and success, however, fell between the bounds. Again, WLL conclude that there is not enough evidence to support the hypothesis that CPG predicts presidential success.

We estimate a model using our approach with an MCMC consisting of four chains, each with a 250,000 iteration burnin, followed by 50,000 iterations and a thinning of 5. The results, then, are based on a total of 40,000 draws from the posterior distribution. The summaries of the posterior are reported in Table 5. Our results are similar to WLL for two

Table 5: 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		

Note: Bayesian error correction model replicating Lebo and O'Green (2022). Results present the median and 95 percent credible region of the parameter estimates from the posterior distribution. LRM = long-run multiplier.

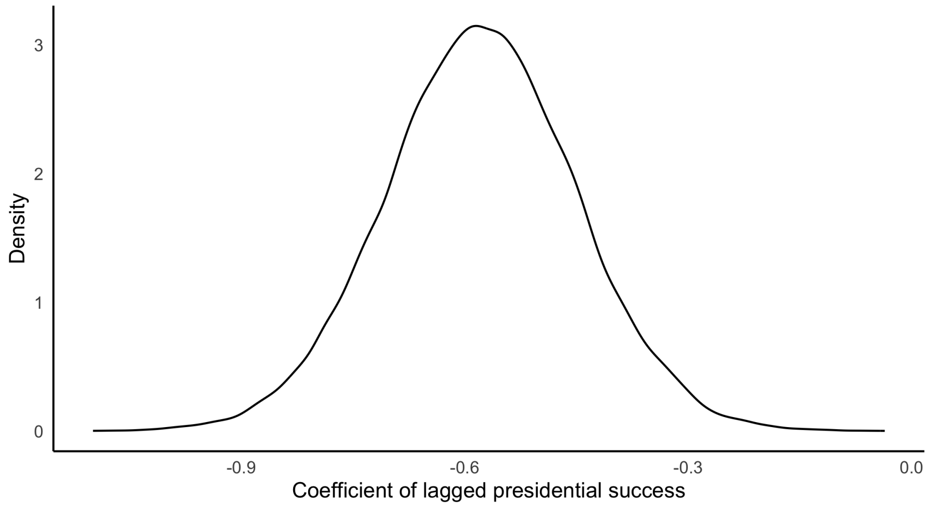
of the three variables. The point estimates of the coefficients and the LRMs reported in the table are almost identical to those reported bay WLL: support for the effect of the share of seats held by the president's party and no evidence for the effect of approval.

Our results differ on the effect of CPG. The 95 percent credible region for the posterior estimate of the LRM excludes zero, indicating that there is an LRR between the variables. In fact, we find that 99.6 percent of the draws from the posterior distribution are greater than zero. In comparison, the posterior of the LRM for the president's party House size is greater than zero in 99.95 percent of the draws, while for presidential approval, this statistic is only 69.3 percent.

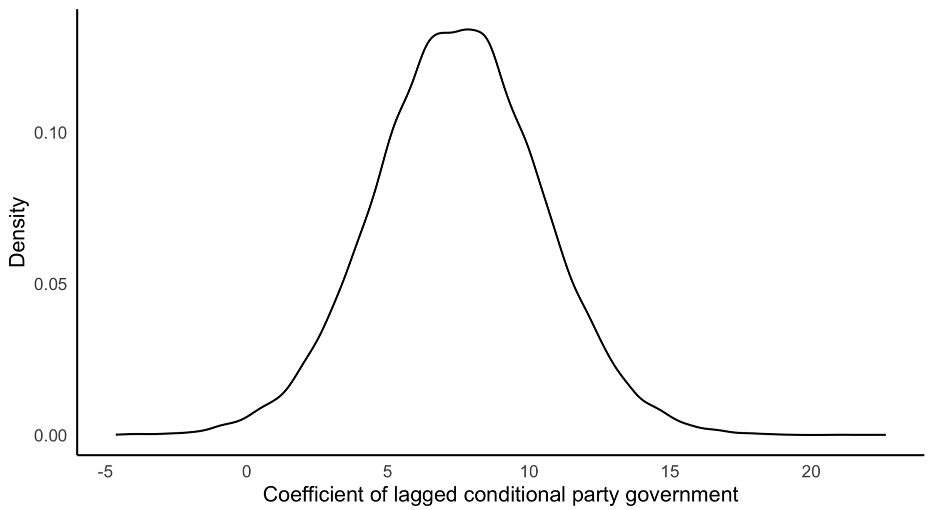
Figure 3 presents the posterior distributions for the coefficients for the lagged dependent variable and CPG, and the LRM for the CPG measure. The figures illustrate the same results as Table 5, but we see the value in presenting the visualization of the posteriors. Every draw from the posterior of the coefficient for the lagged dependent variable is negative, with the maximum value in the draws from the posterior being -0.13. The posterior distribution of the LRM has 99 percent of its mass to the right of zero, indicating that there is a robust LRR between the CPG and presidential success.

Figure 3: Estimated posterior of the coefficient for the lagged values of presidential success and conditional party government, and the long run multiplier for conditional party government.

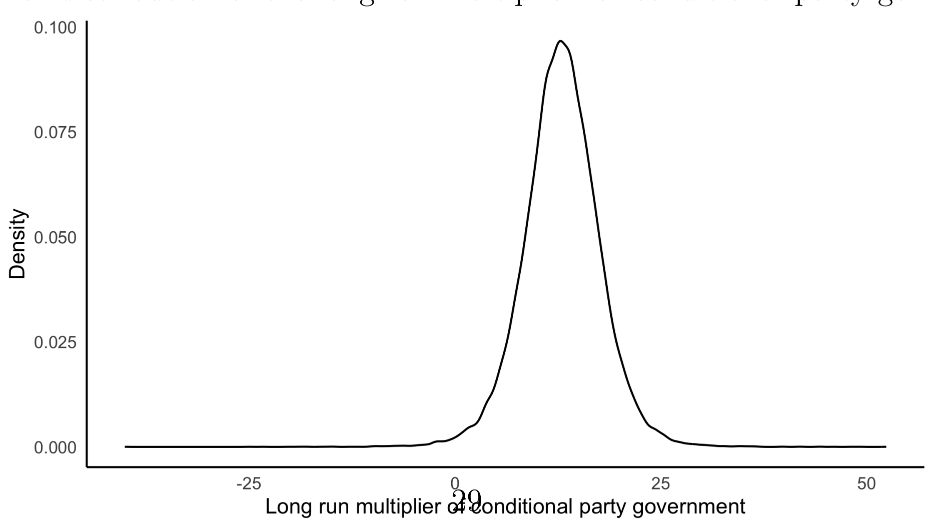
a) Posterior distribution of the coefficient for lagged presidential success.



b) Posterior distribution of the coefficient for lagged conditional party government.



c) Posterior distribution of the long run multiplier for conditional party government.



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 suspect that there are numerous studies that have been started and stopped, and eventually dropped into 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 by 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 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.

We show that a Bayesian approach that directly estimates the LRM from the posterior distribution may address the problem. Using non-informative priors on most parameters, but directly incorporating the limits on the dynamic parameters in the model, we get better coverage in small sample sizes. This framework allows one to readily report uncertainty in the estimates of the LRM and their degree of confidence of a LRR, while also working well with the short time series common with social science data.

While we focus on time series analysis, our approach should apply to other cases where small samples and multiplier effects are present. For example, spatial autoregressive models (Darmofal, 2015), along with their temporal and multiparametric counterparts (Franzese and Hays, 2007; Hays, Kachi, and Franzese, 2010; Cook, Hays, and Franzese, 2023), face some of the same theoretical and empirical constraints regarding stationarity and the potential for standard errors that exceed theoretical limits. Moreover, MCMCs are already frequently used in estimation in closely related fields, such as Markov random fields (Ward and Gleditsch, 2002; Chyzh and Kaiser, 2019) and exponential random graph models (Cranmer and Desmarais, 2011; Desmarais and Cranmer, 2012), with Bayesian extensions existing for each (Box-Steffensmeier, Christenson, and Morgan, 2018; Stundal et al, 2023). Future research could explore the benefits of incorporating semi-informed priors to these and other related areas.

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., Dino P. Christenson, and Jason W. Morgan. 2018. Modeling Unobserved Heterogeneity in Social Networks with the Frailty Exponential Random Graph Model. *Political Analysis* 26(1): 3–19.

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.

Brandt, Patrick T., Michael Colaresi, and John R. Freeman. 2008. The Dynamics of Reciprocity, Accountability, and Credibility. *Journal of Conflict Resolution* 52(3): 343–374.

Brandt, Patrick T., and John R. Freeman. 2006. Advances in Bayesian Time Series Modeling and the Study of Politics: Theory Testing, Forecasting, and Policy Analysis. *Political Analysis* 14(1): 1–36.

Brandt, Patrick T., and John R. Freeman. 2009. Modeling Macro Political Dynamics. *Political Analysis* 17(2): 113–142.

Brandt, Patrick T., and Todd Sandler. 2010. What Do Transnational Terrorists Target? Has It Changed? Are We Safer? *Journal of Conflict Resolution* 54(2): 214–236.

Brandt, Patrick T., and Todd Sandler. 2012. A Bayesian Poisson Vector Autoregression Model. *Political Analysis* 20(3): 292–315.

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

Bürkner, Paul-Christian. 2017. Brms: An R package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software* 80(1): 1–28.

Chan, Joshua, Gary Koop, Dale J. Poirier, and Justin L. Tobias. 2020. *Bayesian Econometric Methods*. 2nd edition. Cambridge University Press.

Chyzh, Olga V., and Mark S. Kaiser. 2019. A Local Structure Graph Model: Modeling Formation of Network Edges as a Function of Other Edges. *Political Analysis* 27(4): 397–414.

Cook, Scott J., Jude C. Hays, and Robert J. Franzese Jr. 2023. STADL Up! The Spatiotemporal Autoregressive Distributed Lag Model for TSCS Data Analysis." *American Political Science Review* 117(1): 59–79.

Cranmer, Skyler J., and Bruce A. Desmarais. 2011. Inferential Network Analysis with Exponential Random Graph Models. *Political Analysis* 19(1): 66–86.

Darmofal, David. 2015. Spatial Analysis for the Social Sciences. Cambridge University Press.

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

Desmarais, Bruce A., and Skyler J. Cranmer. 2012. Statistical Inference for Valued-Edge Networks: The Generalized Exponential Random Graph Model. *PloS One* 7(1): e30136.

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

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai. 2019. Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records. *American Political Science Review* 113(2): 353–371.

Eshima, Shusei, Kosuke Imai, and Tomoya Sasaki. 2024. Keyword-assisted Topic Models. *American Journal of Political Science* 68(2): 730–750.

Fariss, Christopher J. 2014. Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability. *American Political Science Review* 108(2): 297–318.

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

Franzese Jr, Robert J., and Jude C. Hays. 2007. Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data. *Political Analysis* 15(2): 140–164.

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.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*. 2nd edition. Chapman & Hall/CRC.

Gill, Jeff. 2009. *Bayesian Methods: A Social and Behavioral Sciences Approach*. 2nd edition. Chapman & Hall/CRC.

Gill, Jeff. 2012. Bayesian Methods in Political Science: Introduction to the Virtual Issue. *Political Analysis* 20(V3): 1–9. doi:10.1017/S1047198700014273.

Grimmer, Justin. 2010. A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis* 18(1): 1–35.

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

Hays, Jude C., Aya Kachi, and Robert J. Franzese Jr. 2010. A Spatial Model Incorporating Dynamic, Endogenous Network Interdependence: A Political Science Application. *Statistical Methodology* 7(3): 406–428.

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.

Hoff, Peter D., and Michael D. Ward. 2004. Modeling Dependencies in International Relations Networks. *Political Analysis* 12(2): 160–175.

Hollenbach, Florian M., Iavor Bojinov, Shahryar Minhas, Nils W. Metternich, Michael D. Ward, and Alexander Volfovsky. 2021. Multiple Imputation using Gaussian Copulas. *Sociological Methods and Research* 50(3): 1259–1283.

Honaker, James and Gary King. 2010. What to Do about Missing Values in Time-Series Cross-Section Data. *American Journal of Political Science* 54(2): 561–581.

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.

Jackman, Simon. 1994. Measuring Electoral Bias: Australia, 1949–93. *British Journal of Political Science* 24(3): 319–357.

Jackman, Simon. 2001. Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference and Model Checking. *Political Analysis* 9(3): 227–241.

Jordan, Soren and Andrew Q. Philips. 2018a. Cointegration Testing and Dynamic Simulations of Autoregressive Distributed Lag Models *The Stata Journal*. 18(4): 902–923.

Jordan, Soren and Andrew Q. Philips. 2018b. Dynamic Simulation and Testing for Single-equation Cointegrating and Stationary Autoregressive Distributed Lag Models. , 10(2). *The R Journal*. 10(2): doi:10.32614/RJ-2018-07.

Jordan, Soren and Andrew Q. Philips. 2020. Exploring Meaningful Visual Effects and Quantities of Interest from Dynamic Models through dynamac. *Journal of Open Source Software*. 5(54): 2528.

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

King, Gary, and Andrew Gelman. 1991. Systemic Consequences of Incumbency Advantage in U.S. House Elections. *American Journal of Political Science* 35(1): 110–138.

Knutsen, Carl Henrik, Kyle L Marquardt, Brigitte Seim, Michael Coppedge, Amanda B Edgell, Juraj Medzihorsky, Daniel Pemstein, Jan Teorell, John Gerring, Staffan I Lindberg. 2024. Conceptual and Measurement Issues in Assessing Democratic Backsliding. *PS: Political Science and Politics* 57(2): 162–177.

Lebo, Matthew J. and Kraft, Patrick W. 2017. The General Error Correction Model in Practice. *Research and Politics* 4(2): 2053168017713059.

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.

Linzer, Drew A. 2013. Dynamic Bayesian Forecasting of Presidential Elections in the States. *Journal of the American Statistical Association* 108(501): 124–134.

Marriott, John and Paul Newbold. 1998. Bayesian Comparison of ARIMA and Stationary ARMA Models. *International Statistical Review* 66(3): 323–336.

McGinnis, Michael D., and John T. Williams. 1989. Change and Stability in Superpower Rivalry. *American Political Science Review* 83(4): 1101–1123.

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

Minhas, Shahryar, Peter D. Hoff, and Michael D. Ward. 2019. Inferential Approaches for Network Analysis: AMEN for Latent Factor Models. *Political Analysis* 27(2): 208–222.

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.

Montgomery, Jacob M. and Brendan Nyhan. 2010. Bayesian Model Averaging: Theoretical Developments and Practical Applications. *Political Analysis* 18(2): 245–270.

Mueller, Hannes and Christopher Rauh. 2018. Reading Between the Lines: Prediction of Political Violence Using Newspaper Text. *American Political Science Review* 112(2): 358–375.

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.

Nieman, Mark David. 2016. Moments in Time: Temporal Patterns in the Effect of Democracy and Trade on Conflict. *Conflict Management and Peace Science* 33(3): 273–293.

Pang, Xun. 2010. Modeling Heterogeneity and Serial Correlation in Binary Time-Series Cross-Sectional Data: A Bayesian Multilevel Model with AR(p) Errors. *Political Analysis* 18(4): 470–498.

Pang, Xun. 2014. Varying Responses to Common Shocks and Complex Cross-Sectional Dependence: Dynamic Multilevel Modeling with Multifactor Error Structures for Time-Series Cross-Sectional Data. *Political Analysis* 22(4): 464–496.

Park, Jong Hee. 2010. Structural Change in U.S. Presidents' Use of Force. *American Journal of Political Science* 54(3): 766–782.

Park, Jong Hee. 2011. Changepoint Analysis of Binary and Ordinal Probit Models: An Application to Bank Rate Policy Under the Interwar Gold Standard. *Political Analysis* 19(2): 188–204.

Park, Jong Hee. 2012. A Unified Method for Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models. *American Journal of Political Science* 56(4): 1040–1054.

Park, Jong Hee and Sooahn Shin. 2020. Bayesian Methods in Political Science. In *Research Methods in Political Science and International Relations*, eds Curini, Luigi and Robert Franzese. Sage. 895–909.

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 A.M. 2009. Campaign Learning and Vote Determinants. *American Journal of Political Science* 53(2): 445–460.

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

Quinn, Kevin M. 2004. Bayesian Factor Analysis for Mixed Ordinal and Continuous Responses. *Political Analysis* 12(3): 338–353.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science* 58(4): 1064–1082.

Santifort, Charlinda, Todd Sandler, and Patrick T. Brandt. 2013. Terrorist Attack and Target Diversity: Changepoints and Their Drivers. *Journal of Peace Research* 50(3): 75–90.

Schnakenberg, Keith E., and Christopher J. Fariss. 2014. Dynamic Patterns of Human Rights Practices. *Political Science Research and Methods* 2(1): 1–31.

Spirling, Arthur. 2007. Bayesian Approaches for Limited Dependent Variable Change Point Problems. *Political Analysis* 15(4): 387–405.

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

Stundal, Logan, Benjamin E. Bagozzi, John R. Freeman, and Jennifer S. Holmes. 2023. Human Rights Violations in Space: Assessing the External Validity of Machine-Geocoded Versus Human-Geocoded Data. *Political Analysis* 31(1): 81–97.

Thies, Cameron G., and Mark David Nieman. 2017. *Rising Powers and Foreign Policy Revisionism: Understanding BRICS Identity and Behavior through Time.* University of Michigan Press.

van de Schoot, Rens and Milica Miočević. 2020. *Small Sample Size Solutions*. Routledge.

Veen, Duco and Marthe Egberts. 2020. The Importance of Collaboration in Bayesian Analysis with Small Samples. In *Small Sample Size Solutions*, eds van de Schoot, Rens, and Milica Miočević. Routledge. 50–70.

Ward, Michael D. and Kristian Skrede Gleditsch. 2002. Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace. *Political analysis* 10(3): 244–260.

Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. 2010. The Perils of Policy by P-value: Predicting Civil Conflicts. *Journal of Peace Research* 47(4): 363–375.

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.

Western, Bruce. 1998. Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach. *American Journal of Political Science* 42(4): 1233–1259.

Wolak, Jennifer, and David A. M. Peterson. 2020. The Dynamic American Dream. *American Journal of Political Science* 64(4): 968–981.