Political Analysis (2016) 24:69–82
doi:10.1093/pan/mpv035

Equation Balance and Dynamic Political Modeling

Matthew J. Lebo and Taylor Grant
Department of Political Science, Stony Brook University, Stony Brook, NY 11794
e-mail: matthew.lebo@stonybrook.edu (corresponding author)

Edited by Janet Box-Steffensmeier

The papers in this symposium agree on several points. In this article, we sort through some remaining areas of disagreement and discuss some of the practical issues of time series modeling we think deserve further explanation. In particular, we have five points: (1) clarifying our stance on the general error correction model in light of the comments in this issue; (2) clarifying equation balance and discussing how bounded series affects our thinking about stationarity, balance, and modeling choices; (3) answering lingering questions about our Monte Carlo simulations and exploring potential problems in the inferences drawn from long-run multipliers; (4) reviewing and defending fractional integration methods in light of the questions raised in this symposium and elsewhere; and (5) providing a short practical guide to estimating a multivariate autoregressive fractionally integrated moving average model with or without an error correction term.

1 Introduction

The comments of Keele, Linn, and Webb (2016, KL&W hereafter), Freeman (2016), Helgason (2016), and Esarey (2016) on the issues brought up in our paper *Error Correction Methods with Political Time Series* are extremely useful. In particular, this symposium provides some much-needed discussion about relating different types of time series to various modeling strategies. Our main goal in Grant and Lebo (2016, G&L hereafter) is to point out under-appreciated problems with the general error correction model (GECM). While not the authors' intent, many have interpreted *Taking Time Seriously* by DeBoef and Keele (2008) as providing *carte blanche* for when the GECM is appropriate. We see applying and interpreting the GECM as not nearly as simple a task as has been assumed. Our discussions of bounded series, equation balance, and fractional integration methods also sparked useful comments, and we appreciate the opportunity to clarify our thoughts on these issues.

The papers in this symposium agree on several points. Here, we clarify our arguments, sort through some remaining areas of disagreement, and discuss some practical issues for time series modeling. In particular, we (1) clarify our stance on the GECM in light of the comments in this symposium; (2) discuss how bounded series affects our thinking about stationarity, equation balance, and modeling choices; (3) answer lingering questions about our Monte Carlo simulations and explore potential problems in the inferences drawn from long-run multipliers (LRMs); (4) review and defend fractional integration methods; and (5) provide a short practical guide to estimating a multivariate autoregressive fractionally integrated moving average (ARFIMA) model with or without an error correction term.

2 The GECM in Practice

To summarize the main points of G&L: we view the GECM as extremely limited in its applicability—as we say in the abstract: "the model is treated as perfectly flexible when, in fact, the opposite is true." Also, we see the use of the GECM in political science as rife with errors. The GECM may provide

Authors' note: We are grateful to the Editors of *Political Analysis* for the opportunity to respond to the comments in this issue. We also thank Janet Box-Steffensmeier for her help throughout this project. Replication materials are available online as Lebo and Grant (2016).

correct inferences when data are exactly appropriate and researchers are careful in their post-estimation calculations and interpretations. Although we did not investigate every GECM application, at least one of these critical mistakes was made in every paper we did read: estimating unbalanced equations, using a unit-root dependent variable with the wrong critical values, improper interpretation of the ECM and beta parameters with non-unit-root data, or using bounded series with no attention paid to the consequences. Our article re-analyzes only five articles, but in this exchange we did not see a defense of any GECM results published by a political scientist.

Our article emphasizes that researchers have generally not been careful with the GECM and our simulations, in part, demonstrate the consequences of making common errors. The pattern we see in the literature is the misuse of D&K's equation (5):

$$\Delta Y_t = \alpha_0 + \alpha_1^* Y_{t-1} + \beta_0^* \Delta X_t + \beta_1^* X_{t-1} + \epsilon_t. \tag{1}$$

D&K and KL&W explain the equivalence of the ECM to the autoregressive distributed lag (ADL):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t. \tag{2}$$

We recognize the mathematical equivalence between the ADL and the GECM but the estimated parameters of the two are not interchangeable on a one-to-one basis; the equivalence comes through $\alpha_1^* = (\alpha_1 - 1)$, $\beta_0^* = \beta_0$, and $\beta_1^* = \beta_0 + \beta_1$. In the output from the GECM equation, a researcher will often have at least one additional significant estimate despite the lack of any significant temporal dynamics. Table 5 in G&L shows a simple bivariate example where the ECM looks strongly significant even though, as KL&W note, this indicates the *lack* of any long-run relationship between Y and X. The significance of the parameter is deceiving and has generally fooled researchers. In practice, α_1^* is treated as though the data have unit-roots and are cointegrated—reports of strong error correction appear frequently.

Additionally, the GECM adds two ADL parameters together $(\beta_0 + \beta_1)$ while the standard errors are not additive. G&L's table 5 shows the increased risk of Type I errors for X when using the GECM rather than the ADL. Researchers usually skip the post-estimation calculations that may only make their results look weaker and less interesting. Our paper and examples such as our table 5 demonstrate the inferential mistakes made when one simply estimates D&K's equation 5 (equation (1)) and takes the output at face value.

KL&W push us and future researchers to focus on the LRM as the key quantity of interest, and they provide useful clarification on the steps one should take post-estimation. We discuss below why the LRM can also be problematic, but we can note here that Table 10 in G&L shows the LRMs for Shark Attacks and Beef Consumption to be significant predictors of Supreme Court Liberalism and Table 13 has seven of ten significant LRMs for our nonsense series' effects on public mood. Researchers, reviewers, and journal editors should be aware of these possibilities.

KL&W say: "We argue that the results in the Grant and Lebo replications stem from inadequate sample sizes that make it difficult to conclusively use any time series model." We share their worry that researchers are trying to squeeze too much out of small sample sizes—the GECM would have them estimate short- and long-term effects and then compute LRMs. We also note that if Keele and Kelly (2006) are correct that there must be 250 observations for any diagnostic tests to effectively work with a lagged DV, then the use of the GECM with most political time series means we cannot tell whether our models are properly specified. This is yet another serious drawback.

Error correction is an interesting phenomenon and makes for a good story when it is found; for example, macro-level variables that seemingly move in harmony over decades. However, if data do not contain unit-roots, a researcher who claims to have found strong error correction has most likely only found confirmation of some form of stationarity—there may be no effects of the Xs at all. Also, the very concept of equilibrium is different with non-unit-root variables. If both Y and X are mean stationary then, in the long term, they are both returning to their respective means. What would error correction be in that case and why should we be so interested in it? And, if data are bounded, then inferences are muddied further. In all, with stationary series, we doubt the value of a model that differences them, eats up degrees of freedom, potentially confounds diagnostic tests, and opens so many doors for misinterpretation. For these and other reasons, if one is not going to use a pre-filtering approach, we recommend the ADL for stationary (but not fractionally integrated) series.

3 Equation Balance, Bounded Variables, and Modeling Choices

One point of agreement among the papers here is that equation balance is an important and neglected topic. One cannot mix together stationary, unit-root, and fractionally integrated variables in either the GECM or the ADL.[1] Data of interest do not usually all line up as similar in their univariate properties. Thus, balance is hard to come by without transforming some series. Despite general agreement, in practice there are reasons why we might still disagree about whether a particular equation is in or out of balance.

First, a researcher who shares KL&W's concern that fractional integration techniques—including the problems of estimating d—are too unreliable is left to choose between models that difference series by one or not at all. For example, KL&W support using FI techniques when one has confidence that data are FI, but say: "...drawing inferences about the existence and extent of fractional integration is problematic in sample sizes typically seen in political science..." We provide counter-arguments below but, for now, the important point is that this view leads one back to considering the question of integration as a 0/1 dichotomy.

Once locked into a dichotomous choice, authors then consider whether or not their series contain unit-roots. Missteps here are easy if we diagnose the properties of our series in terms of some population instead of the sample in hand. This approach has its proponents; for example, Williams (1992) says: "Classical inference is, as we all know, based on inferring something about a population from a sample of data. In time series, the sample is not random, and the population contains the future as well as the past." This could lead a researcher to consult textbook definitions of stationarity such as KL&W offer: "A weakly stationary time series is one for whom the mean, variance, and covariance are time invariant."

It is at this point that the boundedness of so many political time series poses major problems. For one, a series that has both upper and lower limits will have finite variance and will have some long-term mean to which it will eventually revert.[2] These facts are frequently cited by researchers as evidence that their bounded series cannot have unit-roots and must by definition be stationary. Stationarity tests are often eschewed as authors simply point at the boundedness of their variables.[3] Second, when stationarity tests—especially the Dickey–Fuller test—are administered to bounded data, they are biased toward concluding stationarity (Cavaliere and Xu 2014). These missteps might lead a researcher to mistakenly conclude a bounded series is stationary.

In G&L, we point out the prevalence of bounded political time series—thirteen of thirteen dependent variables in our applied examples—and we investigate some of the statistical problems they create. But we only briefly discuss the issues they pose for diagnosing data.

To elaborate, a researcher using bounded series might accept the argument that bounded series cannot contain unit-roots or she might accept incorrect results from biased stationarity tests. Doing either for all her series would lead to the conclusion that all are stationary. From this, several assumptions would follow, including: (1) the equation is balanced; (2) cointegration is not required for an error-correction specification; and (3) standard normal critical values are appropriate.

[1]Freeman says: "...GL's recommendation to 'set aside' unbalanced equations are a bit overdrawn" and that models can be estimated with unbalanced data so long as nonstandard distributions are used. The recommendation Freeman cites is in specific reference to the cointegration test implied within the GECM. When Y and X are integrated but are not cointegrated with each other, then the $(Y_{t-1} - X_{t-1})$ vector of the GECM will be non-stationary, the equation will be unbalanced, and α_1 will be biased. In such a case, the ECM is not a useful model, although one could still regress ΔY_t on ΔX_t. More generally, we view using unbalanced data as acceptable if one uses a pre-whitening approach or follows the advice cited by Freeman.

[2]For example, if we had 10,000 months of presidential approval data, we would see it revert continually to some stationary mean—a graph on a page might look like it came from a Richter Scale machine. But over the period of data we actually have—650 months at best—the series measures as long memoried, fractionally integrated, or a unit-root (e.g., Lebo and Cassino 2007).

[3]For example, Williams (1992, 230) says: "Presidential Approval also is bounded between zero and one, placing an important theoretical limit on its dynamics." Ehrlich (2007) states: "However, given that tariffs are bounded, unit-roots are not technically possible. Instead, tariffs are probably near-integrated."

In short, a researcher who dismisses fractional methods makes decisions in a 0/1 world and, if she accepts the finite variance argument or faulty test results, she will dismiss the possibility of $I(1)$—all bounded series will be assumed to be not-integrated and therefore all equations that include them are assumed to be balanced.

So, although we all agree that equations should be balanced, a theoretical appeal to the asymptotic properties of time series can lead one to assume that series are stationary and thereby forgo closer examination of equation balance. For example, D&K's 2008 article is meant to apply to stationary data but in their two applied examples the dependent variables are the effective tax rate on labor and Congressional approval, both bounded.[4] Independent variables include the structural unemployment rate and the Index of Consumer Sentiment, respectively—all bounded. Perhaps based on the "bounded series cannot contain unit-roots" argument, stationarity tests are not presented, the series are assumed to be stationary, and the standard normal distribution is used for ECM critical values.

We see the "bounded series cannot contain unit-roots" argument as a wrong turn on the way to properly modeling dynamic relationships. For a stylized example, Fig. 1 presents the *Trade Balance* and *Unemployment Rate* for the planet Caprica. The Trade Balance shown on the left is unbounded—it can theoretically approach positive or negative infinity and the series will eventually exhibit infinite variance. The Unemployment Rate shown on the right, however, has both upper and lower bounds giving it finite variance; that is, it fails the textbook definition for a unit-root. However, aside from the variable names and the scale on the Y-axis, the two series are identical. What is a researcher to do? Should she cite the asymptotic properties of the series and treat them differently?

Our approach is a practical one—to difference or not to difference is the critical empirical choice. "Analysts should deal with the properties of the data sample they have and not make arguments about asymptotics" (Durr 1992). If autocorrelation exists in the data we have, it can affect estimates and requires attention and careful modeling choices.

In the Caprican case, the two series are identical, have identical patterns of autocorrelation, and should be dealt with in the same way to avoid faulty inferences. We advise setting aside the boundedness issue as a decision-maker. Rather, decisions should be made based on rigorous testing of the data in hand using unit-root tests and direct estimates of the fractional integration parameter in (p, d, q) models in addition to careful study of ACF and PACFs *before and after* any type of differencing. Researchers often only consult the residual autocorrelation of the entire model—an unreliable test when including a lagged dependent variable with a limited number of observations.[5] This does not mean that the model's estimates are free from autocorrelation problems. Researchers must be careful to see that their inputs into the model are safe in order to trust the outputs.

Most importantly, we advise researchers to make their fundamental modeling question: "What needs to be done to the data to ensure a trustworthy hypothesis test?" This means being wary of the spurious regression problems in unfiltered time series when they contain unit-roots (Granger and Newbold 1974), are near-integrated (DeBoef and Granato 1997), or are fractionally integrated (Tsay and Chung 2000). This can be avoided if the autocorrelation of each series is dealt with through pre-filtering. This also allows us to take disparate series and reduce them to a common form—stationary deviations from deterministic factors—and leaves us with a balanced equation and reliable estimates.

We also advise researchers to favor the "What needs to be done . . .?" question over the typical "Are my data stationary?" question.[6] KL&W comment on our paper that "They reject the notion

[4]If this year's tax rates are last year's tax rates unless changed by legislation, they essentially contain unit-roots by definition.

[5]This is noted by KL&W in reference to the findings of Keele and Kelly (2006)—in order to "reliably detect autocorrelation in the residuals of regression models with LDVs . . . one needed sample sizes of between 250 and 500 observations before these tests had much power."

[6]A somewhat better question than "are my data stationary?" is "does this series contain a unit-root?" If it does not, there are many possibilities to explore.

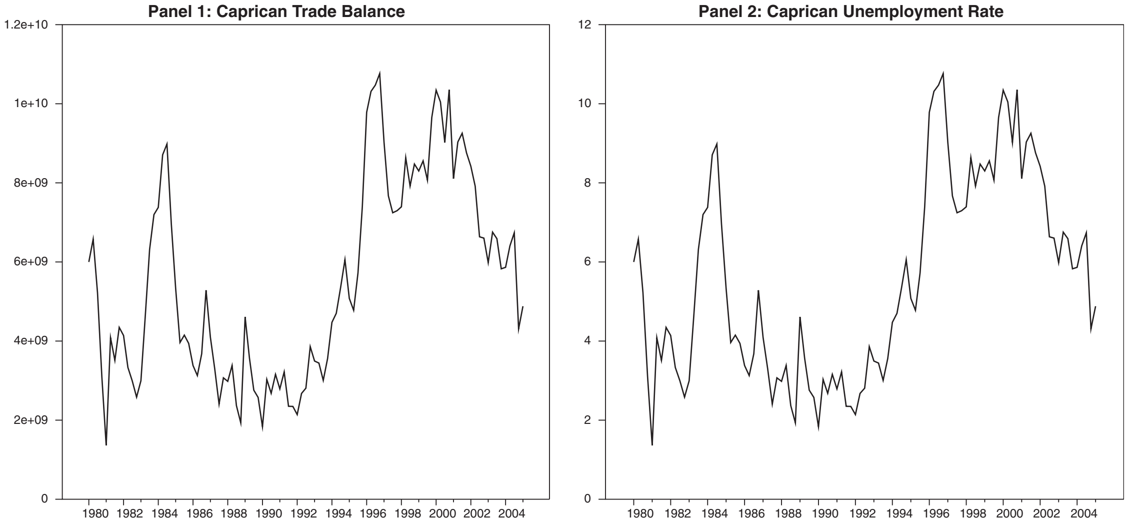

Fig. 1 Trade balance and unemployment rate for Caprica.
Notes. Trade balance is an unbounded variable. The unemployment rate is bounded between 0 and 100.

that most political data is stationary." This leads us down a complicated rabbit hole—stationarity is not an either/or question. There is mean, variance, and covariance stationarity to consider. A mean stationary series could be white noise, autoregressive, near-integrated, fractionally integrated with $d < 0.5$ (mean reverting with finite variance), or fractionally integrated with $d > 0.5$ (mean reverting with infinite variance). Each type poses its own challenges. Left unanswered by the comments in this symposium: if an equation includes two or more of these types of stationary series, is it balanced?

Eschewing testing and filtering because series meet a broad definition of stationarity is a common practice in the literature. *Asking and answering "Is this series stationary?" is simply insufficient in deciding how to proceed with univariate and multivariate modeling*.

4 On Monte Carlos

Questions remain about the simulations we ran in G&L (746 of them) and about those we did not run.[7] The various concerns with our simulations are that we have too many independent variables, our series are too short, our equations are unbalanced, and we are dealing with the wrong quantities of interest. We respond to these critiques next.

4.1 *Overfitting and Balance*

One concern raised is that our models are overfit—that too many variables were included in the Monte Carlo simulations given the short length of the series. While we do include up to five independent variables in some simulations, our results also highlight problems with simple bivariate models. In fact, every set of simulations in G&L includes results from bivariate models and our table 6 provides Type I error rates based solely on bivariate models.

Additionally, our Supplement provides results from bivariate models for near-integrated data in tables G-1 ($T = 60$) and G-6 ($T = 150$) as well as fractionally integrated data in tables H-3 ($T = 60$) and H-8 ($T = 150$). Tables in our Supplement provide multiple types of simulations but in each one

[7]The comments of KL&W, Helgason, and Esarey each provide useful simulations of their own. Still, there are many unexplored questions that further exercises might answer. The provision of code by the authors in this symposium should help spur further research. Simulations found in Tables 1 and 2 were run in Matlab; all replication code and data can be found on the Political Analysis dataverse (Lebo and Grant 2016).

Model 2 (middle row) provides results from perfectly balanced models—($\rho_y = \rho_x$) for near-integrated models or ($d_y = d_x$) for fractionally integrated models. The problems we identify with the GECM still exist in simple bivariate balanced models. Note also that the only table in G&L that uses series bounded by construction is table 3. Thus, our simulations do not allow for the additional possibility that bounds and tests of series with bounds would lead a researcher astray.

4.2 *Series Length and Quantity of Interest*

Two somewhat related critiques of the simulations in G&L are that (1) the series are too short; and (2) we fail to estimate the significance of the LRM—the quantity of interest that researchers should be most interested in.

Our intention was to investigate the GECM as we saw it commonly being used in political science—with a small number of observations, without regard to equation balance or tests of the orders of integration, and using significance tests of the GECM parameters to evaluate the model's key hypotheses. Our approach for choosing replications, simulations, and quantities of interest were informed by these factors. Following the applied papers we read, we did give short shrift to the LRM, and we address that omission here.

Table 1 presents simulations of bivariate GECMs with increasing numbers of observations. As with the data types in G&L's table 6, the specifications were based on at least one statement made in DeBoef and Keele (2008) that the GECM should be favored when data are near-integrated. The

Table 1 Rejection rates for H_0 of GECM (LRM = 0) with near-integrated data

	ρ_x					
***T* = 60**						
ρ_y	0.75	0.80	0.85	0.90	0.95	0.99
0.75	12.9	13.5	13.8	14.2	15.5	17.7
0.80	12.4	13.4	14.0	14.7	17.1	19.7
0.85	12.6	13.4	15.3	16.0	19.4	21.2
0.90	12.2	13.6	15.3	17.9	20.4	23.5
0.95	10.6	12.4	14.8	18.7	21.3	24.2
0.99	10.2	11.6	13.6	16.3	22.4	24.2
***T* = 250**						
ρ_y	0.75	0.80	0.85	0.90	0.95	0.99
0.75	10.1	9.2	9.9	10.3	11.6	11.7
0.80	9.6	9.6	10.3	10.5	12.0	11.9
0.85	9.6	9.2	10.1	10.7	12.0	12.7
0.90	9.7	8.7	9.8	10.7	13.1	14.3
0.95	6.9	7.5	8.4	10.3	13.2	16.0
0.99	4.1	4.3	5.7	7.6	12.3	20.2
***T* = 500**						
ρ_y	0.75	0.80	0.85	0.90	0.95	0.99
0.75	9.7	10.3	10.7	12.5	11.7	12.4
0.80	9.6	10.1	10.8	12.3	12.4	12.4
0.85	9.6	10.0	11.0	12.4	11.3	13.0
0.90	9.3	9.8	10.9	12.5	12.5	13.7
0.95	8.1	8.8	9.6	11.4	13.0	14.7
0.99	3.8	4.5	4.7	6.2	9.6	18.5

Notes. Percentage results based on 1000 simulations. GECM model contains one IV.
GECM: $\Delta Y_t = \alpha_0 + \alpha_1^* Y_{t-1} + \beta_0^* \Delta X_t + \beta_1^* X_{t-1} + \epsilon_t$.

quantity of interest is the LRM, and the standard error is estimated using the variance equation provided by DeBoef and Keele (2008, 192):

$$((1/b^2)\mathrm{Var}(a) + (a^2/b^4)\mathrm{Var}(b) - 2(a/b^3)\mathrm{Cov}(a, b))^{1/2}.$$

Table 1's results do indicate that inferential errors with the GECM's LRM can be reduced, if not eliminated, when series are longer. With sixty observations, there is a significant threat of Type I errors, but with 250 observations the threat of inferential errors is somewhat alleviated, but still too high in many situations. However, in comparison to the results of table 6 in G&L, it is noteworthy that relying on the significance of the LRM rather than the joint hypothesis test of the α_1^* and β_1^* parameters does lead to an increased rate of Type I errors. This occurs because, first, a joint hypothesis test of α_1^* and β_1^* requires passing two significance tests, one of which is two-tailed, whereas the significance of the LRM is based on one parameter's one-tailed test.[8] Second, because the standard error for the LRM is calculated using an equation that provides the asymptotic approximation of the variance, we would expect Type I errors to be biased with smaller samples.[9] This is evident in the discrepancy of Type I error rates when comparing T of 60 and T of 250. With smaller samples, one is likely to overstate the significance of the LRM.

While longer samples will reduce Type I error rates when estimating the GECM with near-integrated data, Table 2 shows that the opposite is true with fractionally integrated data—with FI data, longer series pose greater problems and we are more likely to commit Type I errors. The increasing Type I error rate occurs whether we use a joint hypothesis test on the GECM's α_1^* and β_1^* parameters or if we look only at the estimated LRM using the variance formula provided by DeBoef and Keele (2008). This is yet another reason why proper pre-testing is crucial. Note that these results come from perfectly balanced, bivariate models of unbounded series and, like Table 1, have length up to $T = 500$.

These results are interesting for several reasons. First, because they are FI, the series are likely to appear stationary in simple tests. For instance, with a $d = 0.4$ and $T = 100$, a standard Dickey–Fuller test fails to reject the null of a unit-root 99.9% of the time in 10,000 tests. An augmented Dickey–Fuller with four lags fails to reject 76% of the time. But if we assume that a $d = 0.4$ series is stationary, we are going to greatly overstate the significance of our model and the error rate will increase with the number of observations. Second, this increasing error rate with longer T is the exact opposite of what occurs with near-integrated data. Assuming data are NI when they are FI will cause problems with the GECM. And finally, just as we saw in Table 1, when comparing the Type I errors by the type of hypothesis test used, we find a higher error rate when relying solely on the LRM.

5 Fractional Integration Methods

KL&W (16) say: "GL suggest that much of the data in political science are characterized by fractional integration." Our thoughts here are broken into two parts: fractional integration as a property and ARFIMA modeling as a solution.

First, yes, we think that fractional integration is prevalent in political time series. The theoretical arguments and empirical findings of Box-Steffensmeier and Smith (1996, B-S&S hereafter) are convincing as explanations for the properties of time series created by aggregating individual-level units. As B-S&S explain, Clive Granger's aggregation theorem shows that when individual units are heterogeneous in their degree of autocorrelation, an FI time series results when they are aggregated (Granger and Joyeux 1980). Theoretically, we expect heterogeneity in the electorate in terms of how opinions are remembered among the public (Converse 1964; Key 1966; Zaller 1992). As B-S&S demonstrate, a dichotomous approach to stationarity is a mistake with public opinion data and

[8]As KL&W argue, "The LRM can be statistically significant even if individual terms in the regression model are not."
[9]Esarey (2016) employs the Bewley transformation, which is useful for the LRM but was not used in the articles we read or replicated. When LRMs are used at all, the usual course is to estimate it and its standard errors with the variance equation in DeBoef and Keele (2008, 192). The Bewley method entails estimating a second equation to directly estimate the LRM and its variance. This should cost the GECM points in terms of "ease of use."

Table 2 Rejection rates for H_0 with fractionally integrated data

H_{0A}: LRM = 0

	$d_y = d_x$				
T	0.20	0.30	0.40	0.50	0.60
100	9.5	11.4	14.8	17.7	20.0
200	11.1	15.2	20.2	25.3	26.4
300	12.2	15.8	21.1	24.0	28.9
400	13.7	17.7	22.5	27.0	32.4
500	13.6	18.6	26.1	31.3	34.8

H_{0B}: $\alpha_1^* = 0$ & $\beta_1^* = 0$

	$d_y = d_x$				
T	0.20	0.30	0.40	0.50	0.60
100	5.6	6.9	8.3	10.2	10.8
200	5.3	8.1	12.4	14.6	17.5
300	6.5	8.8	13.5	17.6	19.1
400	7.9	10.5	14.8	18.8	21.3
500	7.0	11.4	17.1	23.3	25.1

Notes. Percentage results based on 1000 simulations. H_{0A}: one-tail significance test. H_{0B}: β_1^* uses two-tail significance test. GECM model contains one IV.
GECM: $\Delta Y_t = \alpha_0 + \alpha_1^* Y_{t-1} + \beta_0^* \Delta X_t + \beta_1^* X_{t-1} + \epsilon_t$.

series like macropartisanship have values of d between 0 and 1. In the 19 years since its publication, *The Dynamics of Aggregate Partisanship* remains largely unchallenged on empirical or theoretical grounds.[10]

Many subsequent empirical studies of opinion data have found series to be fractionally integrated as well (see, e.g., Byers, Davidson, and Peel 2000; Lebo, Walker, and Clarke 2000; Box-Steffensmeier, De Boef and Lin 2004). More widely, it is difficult to think of political time series that are *not* aggregates of individual-level processes (Gil-Alana 2003). Granger's Theorem seems applicable to a wide array of data such as yearly party unity in Congress (Lebo, McGlynn, and Koger 2007) or yearly decisions on the Supreme Court (Lanier 2011; Hendershot et al. 2012).

Closely tied to Granger's Representation Theorem, the expectation of fractional integration is also reasonable when a time series is comprised of the summation of stochastic shocks that survive for varying degrees of time (Parke 1999). Similar to the aggregation of heterogeneous process, if and when a limited number of shocks survive for long periods of time, the underlying aggregate process may be fractionally integrated. Parke (1999) demonstrated this "error duration representation" with the aggregate survival of U.S. firms over a set period of time as justification for the expectation that aggregate unemployment was a fractional process. Grant (2015a) uses the same process to hypothesize that American policy development as well as domestic spending are fractionally integrated processes, and the same theorem could also be used to explain the finding that the size of the Executive Office of the President is a fractionally integrated process (Dickinson and Lebo 2007).

Freeman (2016) states that "... GL are convinced that most political time series are fractional integrated and that unit-roots are rare ... Put simply, we do not yet have the evidence to back up such claims," but he provides neither counterarguments nor citations for why all this evidence is insufficient or in dispute.

[10]For example, scholars have not countered by claiming the electorate is homogeneous in the way voters use past information to inform present opinions. For a public opinion series to be near-integrated, it would mean that all voters have the same strong memory of their past opinions and revert back to their underlying tendency in a homogeneous way.

KL&W are in agreement with us that when data are fractionally integrated and one has good estimates of (p, d, q), ARFIMA models are appropriate and useful. However, KL&W as well as Esarey (2016) have doubts about estimating d. While we might estimate a (p, d, q) model and find a (0, d, 0) model with $0 < d < 1$ to fit well, they would be concerned that the true data-generating process might be something else.[11] In addition, sorting out the exact (p, d, q) model becomes more of a concern as series get shorter—the replications shown in G&L use data that are quite short.

In response, we note that there is more than one way to estimate these parameters. The method that KL&W choose, the time-domain exact maximum likelihood (EML) of Sowell (1992), may be the estimator favored by Stata, but it suffers from a well-known and persistent negative bias (see, e.g., Li and McLeod 1986; Cheung and Diebold 1994; Hauser 1999).

Other estimators fare much better. For (0,d,0) models,[12] semiparametric estimators such as the local Whittle of Robinson (1995), or the log periodogram regression (LPR) model of Geweke and Porter-Hudak (1983), are both unbiased and provide reliable estimates of the d parameter with sample sizes as low as t of 40. If one wishes to estimate d using parametric methods, the frequency domain maximum likelihood (FML) estimator dominates the EML estimator in terms of bias and RMSE. Comparisons of various estimators with various types of (p, d, q) series—primarily with longer series—exist and the benefits of using the FML estimator have been demonstrated with longer time series, most recently by Nielsen and Frederiksen (2005). The ability to reliably estimate the (p, d, q) was investigated again by Grant (2015b), and the FML dominated the EML even with short time series. In particular, the FML is reliable in avoiding Type I errors with respect to FI.[13] Freeman (5) says G&L "paper over the problem of estimation uncertainty" with respect to d. We are concerned about how best to estimate d, but we see the mistakes that might be made estimating d as much smaller than the mistakes made when choosing the wrong multivariate model for fractionally integrated data.

Second, yes, we think that ARFIMA modeling is a useful, flexible, and underused tool. Estimating a (p, d, q) model and then filtering out the noise model is a well-known technique that relies on the logic first set out by Box and Jenkins (1976) and has been shown to be a reliable way to control for autocorrelation when data contain unit-roots or are FI (Hosking 1981; Lebo, Walker, and Clarke 2000; Tsay and Chung 2000). The choice of ARFIMA models can follow theoretical arguments in the literature, investigations of ACF and PACFs, and estimates of (p, d, q) models. In terms of equation balance, ARFIMA methods can impose balance within an equation and allow one to estimate relationships between disparate types of series.

Further, choosing ARFIMA is based on weighing the potential mistakes one could make from using fractional differencing when one should not[14] versus assuming a series requires no differencing when it does. If a variable is truly fractionally integrated but a researcher differences by either one or not at all—essentially rounding to the nearest integer—the ACF of the series used for modeling will be problematic and spurious regressions can follow.[15] There is no evidence we know of that a bigger error is made when one is too quick to use ARFIMA methods rather than too slow. Even if our estimate of d is off the mark by 0.1, this is a benign mistake by comparison.

[11]This is a critical and sticky point to overcome. Despite this symposium and decades of time series work in political science, there is still a lack of trust in diagnostic tools and thus a lack of consensus on how to diagnose the univariate properties of a given time series.

[12]Most ARFIMA work in political science has found that series are well described as (0, d, 0) with $d > 0.5$ (e.g., Byers, Davidson, and Peel 2000). This implies that the simulations in the symposium using (1, d, 0) models are less useful. Of course, KL&W might well argue that a (1, d, 0) series is easily mistaken for a (0, d, 0) series. This is another sticky point. Note also that Box-Steffensmeier and Smith (1998) find some series to be (3, d, 3), but this was due to a coding error in the diagnostics of OX.

[13]For now, Stata only allows use of the inferior EML estimator. Both Robinson's and Geweke's estimators are available in RATS. A wider array of fractional methods may also be found in either R or Matlab. R is open source. Economists using fractional methods are generally quite generous in sharing their Matlab code.

[14]Such as having a complicated ARMA structure as the true DGP but imposing a (0, d, 0) model.

[15]Helgason (2016) asks this critical question in relation to KL&W's claims: "However, in their simulation analysis they do not provide a comparison with the most relevant counterfactual: Namely, whether assuming that data is stationary or integrated while it is actually fractionally integrated is preferable to employing an uncertain estimate of fractional integration and proceeding accordingly."

Again, we see the fundamental question as: What do I need to do to get a trustworthy hypothesis test? We need to be sure our findings are not due to autocorrelation. Otherwise, nothing that comes after that is useful. Nevertheless, more work on effectively diagnosing time series—especially short ones—and understanding what consequences there might be for over-using ARFIMA models would be worthwhile.

6 Building a Multivariate ARFIMA Model—A Brief Practical Guide

Although political scientists have been acquainted with fractional integration since B-S&S's *The Dynamics of Aggregate Partisanship*, FI methods like ARFIMA have not been as widely employed as the GECM. Among the reasons for this are worries about the ability—especially in small samples—to accurately estimate the FI parameter, the complications of estimating the model, and the difficulty of interpreting the results. We dealt with some of these questions above but provide here a brief practical guide for using some FI methods.

An initial question is: how long are your series? Estimating a multivariate FI model with fewer than 50 time-points is perhaps dicey. Statistical packages will still provide estimates of d and will allow fractional differencing, but the error bands around those estimates may get uncomfortably large. On the other hand, with fifty or more time-points you should be able to carefully proceed.[16]

Next, assess the univariate properties of each time series. An essential step here is to simply think about the construction of each series—sometimes this can tell you all you need to know. For example, Segal-Cover Scores (1989) are based on editorials written during a Supreme Court Justice's nomination process. Casillas, Enns, and Wohlfarth (2011) use the Court's yearly median score as a measure of its ideology. If the Court does not change in a year, its score is what it was the previous year. If the Court changes, the new median score is updated. This is unit-root behavior regardless of the fact that Segal-Cover scores are bounded between 0 and 1 and regardless of what test statistics might show. Other examples of unit-roots by construction are cumulative battle deaths—sometimes used in studies of leadership approval (e.g., Mueller 1973)—and yearly tax rates that are usually changed only by legislation (Swank and Steinmo 2002; DeBoef and Keele 2008). It is possible that stationarity tests will disagree, but they might be doubted given what you know about the construction of your data.[17]

Likewise, some time series by their construction may be obviously stationary, particularly those that are first-differences of unit-root series. Examples include GDP growth, the change in the number of Democrats in the House of Representatives, and the daily change in stock indexes. Such series should be investigated for short-term dynamics but are stationary.

Other series may by their construction lead you to expect fractional integration. Following B-S&S's arguments, monthly and quarterly public opinion series are likely cases. This does not mean we should forgo a full array of testing for approval measures and public opinion series such as consumer sentiment, but our initial suspicion should be to expect FI.

The next step is to visually inspect graphs of the series and their corresponding ACFs and partial ACFs. An inspection of the raw series might reveal coding errors, major interventions, or structural breaks that require some attention.[18] For ACFs and PACFs, the second chapter of Enders (2004) is a good source for studying common patterns in series with AR and MA processes. Our Fig. 2 in G&L shows the tell-tale pattern of a fractionally integrated series—an ACF in which the decline is not exponential and correlations are significant at longer lags. KL&W discuss the difficulties in properly diagnosing time series, especially with short samples, but (1) seeing an ACF with a long

[16]This guide assumes data are simply long time series, that is, not indexed by individuals. If data are in panels or repeated cross-sections, multilevel models can be estimated with ARFIMA filtering (Lebo and Weber 2015).

[17]For example, if we use the cumulative number of U.S. casualties in the Vietnam War, we will have a string of zeros in the period before the war begins and a string that remains at the total casualty count after the war has ended. The lengths of these strings will vary based on the scope of our study, but the nature of the series does not change. Stationarity tests will give different results with different time periods.

[18]Freeman (2016, 52) states: "However GL say nothing here about how (if) structural change complicates tests for fractional integration." Young and Lebo (2009) investigate the question of structural breaks in political data and expect them to be quite rare—we often see major movement in time series, but a structural break implies a new and persistent equilibrium level. They also show that estimates of d are not easily biased by such breaks if they do occur.

tail and/or (2) seeing a significant MA parameter in a differenced version of a series are both strong indicators that FI is present.

After those steps, you should have a good sense of what to expect from stationarity and unit-root tests. Testing remains important, and the framework laid out by B-S&S provides a good array of complementary tests—Dickey–Fuller, Variance Ratio, KPSS, and direct estimators of the FI parameter, *d*. As noted above, among the many estimators of *d* the frequency domain estimators offer generally unbiased estimates. The estimates of semiparametric frequency domain estimators such as the local Whittle (Robinson 1995) and LPR (Geweke and Porter-Hudak 1983) are unbiased with (0, *d*, 0) series and can be easily obtained with any statistical software—although Stata does not allow fractional differencing with them. Should the researcher be interested in estimating full (*p*, *d*, *q*) models, the FML estimator will dominate the time domain maximum-likelihood estimator of Sowell (1992) (see Nielsen and Frederiksen 2005; Grant 2015b).[19]

At this point in the process, you should have decided on the appropriate (*p*, *d*, *q*) noise model for each variable. If all of the variables are similarly stationary and not fractionally integrated, then the ADL model is appropriate. If all are unbounded and contain unit-roots, then one could test for cointegration and, if found, use the GECM. But if the variables are of different orders of integration, then models like the ADL and GECM cannot be used since they will be out of balance.

As mentioned, FI methods allow us to create a balanced equation from dissimilar data. By filtering each series by its own (*p*, *d*, *q*) noise model, the residuals of each can be rendered (0, 0, 0) so that you can investigate how *X*'s deviations from its own time-dependent patterns affect *Y*'s deviations from its own time-dependent patterns. Using ARMA, ARIMA, or ARFIMA filters for each variable as appropriate, every series can be reduced to a common and balanced level of integration—0. After filtering the series that require it—that is, not those rare ones that began as white noise—we recommend looking again at graphs of the filtered series and their ACFs and PACFs. No significant autocorrelations should remain. If any do, it is a sign that the series has been filtered incorrectly. For example, applying whole-differencing to a fractionally integrated series will create an over-differenced series likely to contain a significant MA parameter (Dickinson and Lebo 2007).

Next, with pre-whitened versions of each series, you can estimate a regression for short-term effects. Since all the series have had their deterministic components filtered out, all are now (0, 0, 0) and the equation will be balanced—the dirty bathwater has been carefully removed and the regression will yield trustworthy estimates (Hosking 1981).

Our replication and exercises in G&L with data from Casillas, Enns, and Wohlfarth (2011) are a good demonstration of these steps. Although the equation is initially unbalanced (see table C.5), fractionally differencing each series by its estimated *d* value makes for a balanced equation and trustworthy hypothesis tests. Variables that we expect to affect Supreme Court liberalism—Public Mood and Segal-Cover scores—prove to do so in table 11. Pre-whitening has *not* white-washed the data clean of interesting relationships.[20] On the other hand, table C.6 of the G&L Supplement shows that using FI methods fails to turn up the nonsense impacts of Sharks, Tornadoes, and Beef Consumption seen in the GECM's LRMs in table 10. Critics may point out that explaining plainly the meaning of a coefficient with filtered variables is less straightforward, *but we argue that confidence in the results of our hypothesis tests should trump ease of interpretation*.

Testing for error correction is also possible within the FI framework, and we suggest following the three-step fractional cointegration approach demonstrated in Clarke and Lebo (2003) and followed by Helgason (2016) (see also Dueker and Startz 1998). The use of a fractional error correction mechanism (FECM) marries the logic of two-step Engle and Granger (1987)

[19]Because Stata only offers the time domain parametric method, we recommend using either R or Matlab for work on fractional integration.

[20]This is, of course, anecdotal evidence. Helgason's (2016) simulations are especially useful for simulating relationships between fractionally integrated variables including fractional cointegration. Investigating how various models prevent Type II errors is important, but done rarely in this literature. Our quibble with those simulations would be that the data are created to be unbounded and perfectly balanced, giving the GECM a better chance for success that it would have with real-world data.

cointegration testing with fractional differencing and, by relaxing strict assumptions, is quite flexible.[21] Assuming they are of similar orders of integration, the first step—called the cointegrating regression—regresses Y on X just as in E&G's first step. In the second step, the residuals of the cointegrating regression are tested and, if found to have a lower level of (fractional) integration, there is evidence of error correction (Box-Steffensmeier and Tomlinson 2000). These residuals may still be autocorrelated and may require fractional differencing in order to have trustworthy hypothesis tests in the next step.

In the third step, the filtered version of Y is regressed on the filtered version of X and on the filtered and lagged FECM; that is, we estimate $\Delta^{d_Y} Y_t = \alpha_0 + \alpha_1 \Delta^{d_{\mathrm{ECM}}} ECM_{t-1} + \beta_1 \Delta^{d_X} X_t + \epsilon_t$. Error correction implies a very close relationship and may not appear even when X has short-term effects on Y. For example, the data in G&L's table 11 failed to show signs of fractional cointegration in the cointegrating regression and an FECM was not justified in the final model. An applied example that includes a significant FECM is Lebo, McGlynn, and Koger (2007), who estimate yearly Democratic Unity from 1789 to 2000 as (0, 0.69, 0), Republican Unity as (0, 0.78, 0), and Democratic Size as (0, 0.75, 0). Regressing Democratic Unity on the latter two series creates an ECM vector estimated as (0, 0.34, 0). The dependent variable, the independent variables, and the FECM lagged one year are each differenced by their own value of d prior to the final regression.

Post-estimation, one should again check diagnostics such as the Durbin–Watson statistic. Bollerslev et al. (2013) demonstrate impulse response functions with a fractionally cointegrated VAR, and these can be helpful in tracking the long-term impacts of variables on each other. With good diagnostics, one can have confidence in the results of the hypothesis tests for short- and long-term effects, the latter through the FECM.

When comparing FECM and GECM methods, worries about FI's interpretability should be weighed against the findings in G&L of the many factors that can affect the estimation and interpretation of the GECM's parameters. In the GECM, the estimation of α_1 is affected by the number of independent variables, the boundedness of the variables, and the extent of autocorrelation in the variables. Further, the interpretation of α_1 is affected by the researcher's diligence in correctly performing the complicated post-estimation calculations. Practitioners may have an easy time running the GECM—in a single line of Stata code, for example—but it is not easy to arrive at the correctly interpreted and unbiased rate of error correction. In part, this explains why we sometimes find authors explaining error correction rates above 100%. Except in the rare case where one uses the GECM with unbounded and cointegrated unit-root series, the model's α_1 does not simply tell us the rate of error correction. And, as we show in G&L, mistakes for the β parameters are also easy to make.

7 Conclusion

In sum, we remain skeptical that the GECM is a reliable model except in the very rare case where one has unbounded unit-root variables that are cointegrated with each other. Unbalanced equations are a key problem in the large body of work that has cited D&K. With unbalanced equations, fractionally integrated data, or bounded time series, the GECM is prone to spurious results. With stationary data, the estimation of an error correction rate provides little insight into the relationship between the variables. And interpretation problems appear at many points in the process.

Overall, readers should be wary of strong findings coming from the GECM. The model is just not as simple as it seems and not nearly as flexible as has been assumed. Although estimated with a single line of code, it is actually extremely difficult to discern just what the results mean. And, often, researchers have taken raw output with impressive t-statistics at face value, misunderstood its meaning, and published seemingly compelling research.

Use of the GECM has certainly outpaced that of fractional integration methods. Key reasons for this are misperceptions about the flexibility, ease of use, and ease of interpretation for each. Since the FI approach can take disparate series and filter them to impose balance, and since the GECM

[21] Engle and Granger's method requires that the initial series must be unit-roots and that the residuals of a cointegrating regression must be level-stationary.

runs into problems when it squeezes such series into the same equation, it is the FI approach that is far more flexible. Also, ease of use is not only about the ability to run a model—it also includes the ability to avoid misuse of it, to properly interpret the output, and to have confidence that the inferences one makes from it are accurate. In these respects, we think FI methods—when appropriate—dominate the GECM, and we hope that this symposium sparks new interest in them. Failing that, we hope future GECM users pay close attention to the issues raised here and are more careful in their use of the model.

Conflict of interest statement. None declared.

References

Bollerslev, Tim, Daniela Osterrieder, Natalia Sizova, and George Tauchen. 2013. Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics* 108(2):409–24.

Box, George E. P., and Gwilym M. Jenkins. 1976. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Box-Steffensmeier, Janet M., and Andrew R. Tomlinson. 2000. Fractional integration methods in political science. *Electoral Studies* 19(1):63–76.

Box-Steffensmeier, Janet M., and Renée M. Smith. 1996. The dynamics of aggregate partisanship. *American Political Science Review* 90:567–80.

Box-Steffensmeier, Janet M., and Renee M. Smith. 1998. Investigating political dynamics using fractional integration methods. *American Journal of Political Science* 42(2):661–89.

Box-Steffensmeier, Janet M., Suzanna De Boef, and Tse-Min Lin. 2004. The dynamics of the partisan gender gap. *American Political Science Review* 98:515–28.

Byers, David, James Davidson, and David Peel. 2000. The dynamics of aggregate political popularity: Evidence from eight countries. *Electoral Studies* 19(1):49–62.

Casillas, Christopher J., Peter K. Enns, and Patrick C. Wohlfarth. 2011. How public opinion constrains the US Supreme Court. *American Journal of Political Science* 55(1):74–88.

Cavaliere, Giuseppe, and Fang Xu. 2014. Testing for unit roots in bounded time series. *Journal of Econometrics* 178(2):259–72.

Cheung, Yin-Wong, and Francis X. Diebold. 1994. On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *Journal of Econometrics* 62(2):301–16.

Clarke, Harold D., and Matthew Lebo. 2003. Fractional (co)integration and governing party support in Britain. *British Journal of Political Science* 33(2):283–301.

Converse, Philip E. 1964. The nature of belief systems in mass publics. In *Ideology and discontent*, ed. David E. Apter. Ann Arbor: University of Michigan Press.

DeBoef, Suzanna, and Jim Granato. 1997. Near-integrated data and the analysis of political relationship. *American Journal of Political Science* 41(2):619–40.

DeBoef, Suzanne, and Luke Keele. 2008. Taking time seriously. *American Journal of Political Science* 52(1):184–200.

Dickinson, Matthew J., and Matthew J. Lebo. 2007. Reexamining the growth of the institutional presidency, 1940–2000. *Journal of Politics* 69(1):206–19.

Dueker, Michael, and Richard Startz. 1998. Maximum-likelihood estimation of fractional cointegration with an application to US and Canadian bond rates. *Review of Economics and Statistics* 80(3):420–26.

Durr, Robert H. 1992. An essay on cointegration and error correction models. *Political Analysis* 4(1):185–228.

Ehrlich, Sean D. 2007. Access to protection: Domestic institutions and trade policy in democracies. *International Organization* 61(3):571–605.

Enders, Walter. 2004. *Applied econometric time series*. 2nd ed. New York: Wiley.

Engle, Robert F., and C. W. J. Granger. 1987. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55:251–76.

Esarey, Justin. 2016. Fractionally integrated data and the autodistributed lag model: Results from a simulation study. *Political Analysis* 24:42–49.

Freeman, John. 2016. Progress in the study of nonstationary political time Series: A Comment. *Political Analysis* 24:50–58.

Geweke, John, and Susan Porter-Hudak. 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4(4):221–38.

Gil-Alana, Luis A. 2003. Testing of fractional cointegration in macroeconomic time series. *Oxford Bulletin of Economics and Statistics* 65(4):517–29.

Granger, Clive W. J., and Paul Newbold. 1974. Spurious regressions in econometrics. *Journal of Econometrics* 26:1045–66.

Granger, Clive W. J., and Roselyne Joyeux. 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1(1):15–29.

Grant, Taylor. 2015a. Error duration models: A new justification for fractional integration in political time series. PhD thesis. Stony Brook University.

———. 2015b. Fractional integration in short samples: Parametric versus semiparametric methods. PhD thesis. Stony Brook University.

Grant, Taylor, and Matthew J. Lebo. 2016. Error correction methods with political time series. *Political Analysis* 24:3–30.

Hauser, Michael A. 1999. Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study. *Journal of Statistical Planning and Inference* 80(1):229–55.

Helgason, Agnar Freyr. 2016. Fractional integration methods and short time series: Evidence from a simulation study. *Political Analysis* 24:59–68.

Hendershot, Marcus E., Mark S. Hurwitz, Drew Noble Lanier, and Richard L. Pacelle. 2012. Dissensual decision making: Revisiting the demise of consensual norms within the US Supreme Court. *Political Research Quarterly* 66(2):467–81.

Hosking, Jonathan R. M. 1981. Fractional differencing. *Biometrika* 68(1):165–76.

Keele, Luke, and Nathan J. Kelly. 2006. Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political Analysis* 14(2):186–205.

Keele, Luke, Suzanna Linn, and Clayton Webb, 2016. Treating time with all due seriousness. *Political Analysis* 24:31–41.

Key, V. O. Jr. 1966. *The responsible electorate*. Cambridge, MA: Harvard University Press.

Lanier, Drew N. 2011. Acclimation effects and the Chief Justice: The influence of tenure and role on the decisional behavior of the Court's leader, 1888–2007. American Politics Research 39(4):682–723.

Lebo, Matthew J., Adam J. McGlynn, and Greg Koger. 2007. Strategic party government: Party influence in Congress, 1789–2000. *American Journal of Political Science* 51(3):464–81.

Lebo, Matthew J., and Christopher Weber. 2015. An effective approach to the repeated cross-sectional design. *American Journal of Political Science* 59:242–58.

Lebo, Matthew J., and Daniel Cassino. 2007. The aggregated consequences of motivated reasoning and the dynamics of partisan presidential approval. *Political Psychology* 28(6):719–46.

Lebo, Matthew J., and Taylor Grant. 2016. Replication data for: Equation balance and dynamic political modeling. Harvard Dataverse, V1. http://dx.doi.org/10.7910/DVN/SACLMT.

Lebo, Matthew J., Robert W. Walker, and Harold D. Clarke. 2000. You must remember this: Dealing with long memory in political analyses. *Electoral Studies* 19(1):31–48.

Li, W. K., and A. Ian McLeod. 1986. Fractional time series modelling. *Biometrika* 73(1):217–21.

Mueller, John. 1973. *War, presidents and public opinion*. New York: Wiley.

Nielsen, Morten Ørregaard, and Per Houmann Frederiksen. 2005. Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews* 24(4):405–43.

Parke, William R. 1999. What is fractional integration? *Review of Economics & Statistics* 81(4):632–38.

Robinson, Peter M. 1995. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23(5):1630–61.

Sowell, Fallaw. 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* 53(1):165–88.

Swank, Duane, and Sven Steinmo. 2002. The new political economy of taxation in advanced capitalist democracies. *American Journal of Political Science* 46(3):642–655.

Tsay, Wen-Jay, and Ching-Fan Chung. 2000. The spurious regression of fractional integrated processes. *Journal of Econometrics* 96:155–182.

Williams, John T. 1992. What goes around comes around: Unit root tests and cointegration. *Political Analysis* 4:229–35.

Young, Everett, and Matthew J. Lebo. 2009. Long memory methods and structural breaks in public opinion time series: A reply to pickup. *Journal of Elections, Public Opinion and Parties* 19(1):117–24.

Zaller, John R. 1992. *The nature and origins of mass opinion*. New York: Cambridge University Press.