Political Analysis (2016) 24:3–30
doi:10.1093/pan/mpv027

Error Correction Methods with Political Time Series

Taylor Grant
Stony Brook University
e-mail: Taylor.Grant@stonybrook.edu

Matthew J. Lebo
Stony Brook University
e-mail: matthew.lebo@stonybrook.edu (corresponding author)

Edited by Janet Box-Steffensmeier

While traditionally considered for non-stationary and cointegrated data, DeBoef and Keele suggest applying a General Error Correction Model (GECM) to stationary data with or without cointegration. The GECM has since become extremely popular in political science but practitioners have confused essential points. For one, the model is treated as perfectly flexible when, in fact, the opposite is true. Time series of various orders of integration—stationary, non-stationary, explosive, near- and fractionally integrated—should not be analyzed together but researchers consistently make this mistake. That is, without *equation balance* the model is misspecified and hypothesis tests and long-run-multipliers are unreliable. Another problem is that the error correction term's sampling distribution moves dramatically depending upon the order of integration, sample size, number of covariates, and the *boundedness* of Y_t. This means that practitioners are likely to overstate evidence of error correction, especially when using a traditional t-test. We evaluate common GECM practices with six types of data, 746 simulations, and five paper replications.

1 Introduction

Nearly ninety years after Yule (1926) published "Why do we sometimes get nonsense correlations between time-series?" political applications are still plagued by spurious findings. Researchers often favor models that are easier to both implement and interpret and, when data are scarce, the desire for simplicity grows. Nevertheless, some models are particularly prone to give unreliable results.

Given the peculiarities of political data, one such model is the general (unrestricted) error correction model (GECM), first developed in econometrics (Davidson et al. 1978; Hendry, Srba, and Yeo 1978) and introduced to political science over two decades ago (Beck 1992; Durr 1993; Ostrom and Smith 1993).[1] More recently, DeBoef and Keele (2008, D&K) re-introduced the GECM as an equivalent model to the autoregressive distributed lag (ADL) in an *AJPS* workshop piece, "Taking Time Seriously." D&K describe several promising aspects for stationary data: cointegration is unnecessary to look for error correction and one can estimate both short- and long-term impacts of covariates. Among political scientists, the method has since become extremely popular, with D&K used as the go-to source. Google Scholar shows the paper cited 68 times in 2013, with many applications appearing in top journals.[2]

Authors' note: Previous versions of this article were presented at the 2014 meeting of the Midwest Political Science Association, the 2014 meeting of the European Political Science Association, and the 2014 PolMeth conference. We are grateful for helpful comments from Neal Beck, Matthew Blackwell, Janet Box-Steffensmeier, Patrick Brandt, Justin Esarey, Stanley Feldman, Paul Kellstedt, Suzanna Linn, Vera Troeger, and seminar participants at Rice University and Stony Brook University. Replication materials are available online as Grant and Lebo (2016). Supplementary materials for this article are available on the *Political Analysis* Web site and at: https://sites.google.com/a/stonybrook.edu/matthew-lebo/about/resume/papers.

[1]See Banerjee et al. (1993) for the clearest exposition on the topic.

[2]As of October 2014 various versions of the article have been cited 335 times. Recent examples of applied GECM papers include: Kono (2008); Jennings and John (2009); Kayser (2009); Ramirez (2009); Kelly and Enns (2010); Layman et al.

© The Author 2016. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oup.com

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

However, there are several problems with the GECM—some known and some explored here. For one, there are many types of political time series—integrated, near-integrated, fractionally integrated, auto-regressive, and explosive—and researchers need to pay very close attention to the properties of their series. Each presents its own challenges in the GECM. Second, practitioners often fail to ensure that their equations are *balanced*. The order of integration needs to be consistent across all series in a model. Mixing together series of various orders of integration will mean a model is misspecified. Third, although the autoregressive distributed lag (ADL) model is algebraically equivalent to the GECM, the reorganization of parameters is not benign and easily leads to misinterpretation. Fourth, with stationary data, the ECM's coefficient is misinterpreted so that mean reverting variables with little connection to each other are claimed to be in equilibrium relationships. The interpretation of parameters should change across data types. Fifth, with unit-root data, the GECM's key hypothesis test is misunderstood: political science applications have incorrectly used a standard *t*-test for the ECM's coefficient. Sixth, using bounded series in the GECM has unexplored consequences. Given these and other issues, the GECM should only be used in rare instances, yet it is applied frequently and haphazardly. As a result, serious errors exist in the growing body of work that has relied upon the model.

We continue below with a time-series primer, followed by a discussion of the GECM's performance with six types of dependent variables. Simulations demonstrate the issues with each case—typically an alarming rate of Type I errors but sometimes nonsensical inferences as well, e.g., error correction rates that can be misinterpreted to be above 100%. We also outline the problems presented by bounded political time series. Last, we replicate two recent GECM studies: Casillas, Enns, and Wohlfarth (2011) and Ura and Ellis (2012).[3] In each, standard practices greatly overstate the strength of relationships, especially concerning error correction. Without an appreciation of the method's limitations our understanding of dynamic political relationships will continue to be undermined.

2 Unit-Roots, Stationarity, and Error Correction

The univariate properties of time series Y_t can be described as

$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-L)^d Y_t = (1+\sum_{i=1}^{q}\theta_i L^i)\epsilon_t, \tag{1}$$

where p refers to auto-regressive parameters, q refers to moving average parameters, and d is the (fractional) integration parameter. These can be summarized in a (p, d, q) notation where (0, 0, 0) is a white-noise, auto-correlation-free process (Box and Jenkins 1976; Granger and Joyeux 1980; Hosking 1981; Enders 2004). L is the lag operator such that $LY_t = Y_{t-1}$.

The parameter d represents the memory of the series. If $d=0$ the series is weakly stationary and has only short memory—it will tend towards a constant mean, and has finite variance and constant covariance.[4] Autoregressive (AR) and moving average (MA) parameters may still exist where shocks will persist for finite periods as the series reverts back to its mean.[5] If $d=1$, the series contains a unit-root (also known as integrated, $I(1)$, perfect-memoried, or a random-walk) with a non-stationary mean, variance, and covariances. It can wander in any direction with no expectation of returning to a long-term mean.

(2010); Ura and Wohlfarth (2010); Casillas, Enns, and Wohlfarth (2011); Faricy (2011); Sanchez Urribarri et al. (2011); Rickard (2012); Ura and Ellis (2012); Volscho and Kelly (2012); Büthe and Milner (2014); Enns (2014); Enns et al. (2014); Ura (2014).

[3]Three additional replications are in the Supplement: Sanchez Urribarri et al. (2011), Kelly and Enns (2010), and Volscho and Kelly (2012). See Appendices E.1, E.2, and E.3 for respective details. All replication code and data can be found on the *Political Analysis* dataverse (Grant and Lebo 2016).

[4]Enders (2004, 54) describes a weakly stationary series as: mean constant ($E(Y_t) = E(Y_{t-s}) = \mu$), variance constant ($E[(Y_t-\mu)^2] = E[(Y_{t-s}-\mu)^2] = \sigma_y^2$) or ($var(Y_t) = var(Y_{t-s}) = \sigma_y^2$), and covariance constant ($E[(Y_t-\mu)(Y_{t-s}-\mu)] = E[(Y_{t-j}-\mu)(Y_{t-j-s}-\mu)] = \gamma_s$) or ($cov(Y_t, Y_{t-s}) = cov(Y_{t-j}, Y_{t-j-s}) = \gamma_s$).

[5]Near-integrated series have AR processes of almost 1 and are "long-memoried." The effects of shocks decrease at an exponential rate and the series are eventually stationary as well (DeBoef and Granato 1997).

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

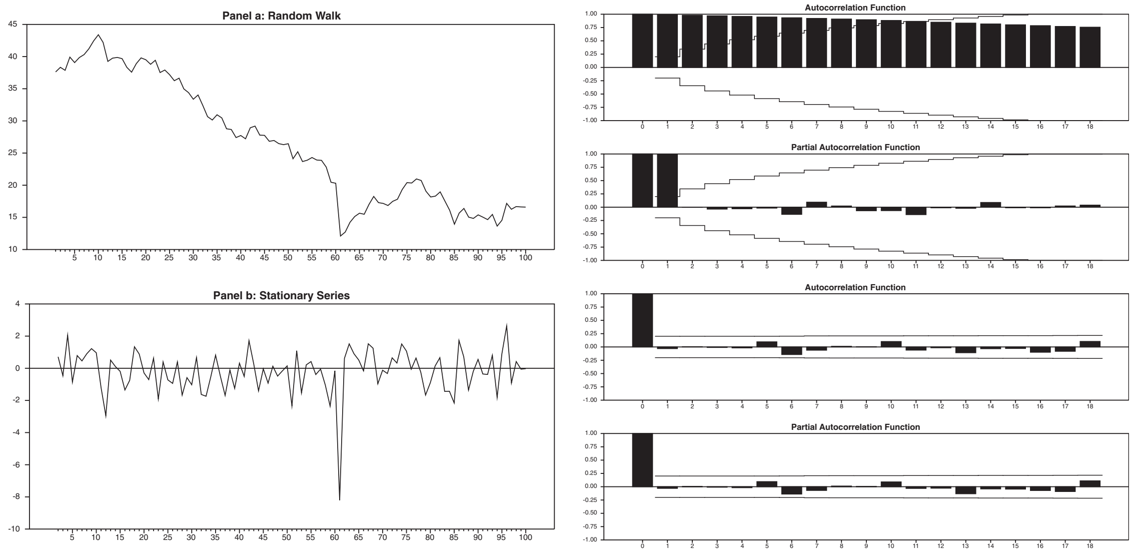

Fig. 1 An integrated series pre- and post-differencing with ACF and PACF.
Note. Spikes outside the solid lines on the auto-correlation and partial autocorrelation function indicate problematic correlations (not including the first spike at $t = 0$).

If we relax the assumption that the order of integration must be an integer, but can instead fall anywhere on a real number line, a series is considered fractionally integrated. If a series is $-1/2 < d < 1/2$, the process is invertible and possesses a linear representation. For $0 < d < 1$, the process is said to have long memory—it holds a mix of characteristics with long—not short or perfect—memory (Granger and Joyeux 1980; Beran 1994; Box-Steffensmeier and Smith 1996). If $1/2 < d < 1$, the series is variance and covariance non-stationary; however, it is still mean reverting (Baillie 1996). Fractional integration is widespread in political data (Box-Steffensmeier and Smith 1996; Box-Steffensmeier and Tomlinson 2000; Lebo, Walker, and Clarke 2000).

As any of the (p, d, q) parameters deviate from zero, auto-correlation threatens hypothesis tests. Extensive work has concentrated on identifying auto-correlation and filtering to account for it (e.g., Granger and Newbold 1974; Box and Jenkins 1976; Granger and Joyeux 1980; Clarke and Stewart 1994; Box-Steffensmeier and Smith 1996, 1998; Lebo, Walker, and Clarke 2000; Tsay and Chung 2000; Clarke and Lebo 2003). This "pre-whitening" approach (1) identifies how a series depends upon its past values, then (2) models this behavior, and (3) uses the "white-noise" residuals so that data are *i.i.d.* and inferences are trustworthy. Pre-whitening approaches have developed from Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) (Box and Pierce 1970; Box and Jenkins 1976) to Autoregressive Fractionally Integrated Moving Average (ARFIMA) models (Granger and Joyeux 1980; Hosking 1981). Researchers may be concerned about what is lost with pre-whitening—the common metaphor asks if one throws out the baby (interesting variation) with the bathwater (auto-correlation). But white noise does *not* mean that all useful information has been removed, leaving only random noise. Rather, it means that the series' dependence on its own history has been filtered out and what is left—*seemingly* white noise—might be explained by independent variables. Thus, pre-whitening ensures that data hold the *i.i.d.* property.

The simplest example of pre-whitening begins with a unit-root, (0, 1, 0): $Y_t = 1 * Y_{t-1} + \epsilon_t$. Figure 1, top-left, shows a unit-root series with its auto-correlation (ACF) and partial auto-correlation (PACF) plots to the right—significant spikes beyond the 0^{th} lag are deviations from *i.i.d.* and will cause problems.[6] First-differencing subtracts Y_{t-1} from each side to create the white noise

[6]Techniques such as estimating and differencing by d lose some precision with smaller samples. But with simple diagnostic tools like ACFs and PACFs, researchers can determine if they have properly modeled out auto-correlation. Fractional integration filters can also be approximated using several AR and MA parameters (Hosking 1981).

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

series, $Y_t^* = \epsilon_t$ with $\epsilon_t \sim N(0, \sigma^2)$ (bottom-left). As shown in Y_t^*'s ACF and PACF, it is free of auto-correlation. Regressions involving it will provide unbiased, efficient, and consistent estimates. Other filters can make other types of data safe as well: Box and Pierce (1970) use ARMA (p, q) filters for auto-regressive and moving average processes, Box-Jenkins (1976) adapted it to unit-roots (ARIMA models), and Hosking (1981) introduced fractional differencing (ARFIMA) to filter a (p, d, q) model into white noise, (0, 0, 0).

Whether (p, 0, q), (p, 1, q), or (p, d, q), the goal is the same: create white noise residuals for *all* variables to protect inferences (Granger and Newbold 1974; Clarke and Stewart 1994; Tsay and Chung 2000; Clarke and Lebo 2003). Granger and Newbold (1974) find significant (0.05 level) relationships about 75% of the time between two randomly generated unit-roots; Lebo, Walker, and Clarke (2000) find similar problems with fractionally integrated data, and DeBoef and Granato (1997) do so with near-integrated data. In every case, pre-whitening solves these problems and minimizes Type I errors.

The inferential threats of auto-correlation are not debated, nor do researchers argue about whether pre-whitening works. No effort that we know of has sought to explain why under-differencing or over-differencing are not really problems. Yet, pre-whitening is often skipped. One worry is that information is lost by filtering—the "babies and bathwater" argument. Also, studying a series in differences means we cannot study long-term relationships (Beck 1992; Bannerjee et al. 1993).

The desire to study both short- and long-term relationships motivates error correction methods. Classic cointegration posits equilibrium relationships so that shocks that separate series are short-lived and error correction mechanisms (ECMs) measure the rate of re-equilibration (Engle and Granger 1987). The two-step ECM approach begins with $I(1)$ variables and tests whether they are cointegrated—that is, is a linear combination of them $I(0)$? In the first step, Y_t is regressed on covariates and residuals are tested for stationarity. If the residuals are $I(0)$, then a second step uses the lagged residuals as the ECM in a differenced model:

$$\Delta Y_t = \beta_0 + \beta_1 \Delta X_t + \alpha_1 ECM_{t-1} + \epsilon_t. \tag{2}$$

Both short- (β_1) and long-run (α_1) effects are captured. The separation of Y_{t-1} and X_{t-1} (errors) are corrected at time t as the series return to equilibrium. Many Xs may affect Y but few will prove to be in an equilibrium relationship.[7]

Many alternative models exist, including one-step GECMs that skip testing specifically for cointegration. Single-equation GECMs have quickly become the most popular ECM technique among political scientists who have seen it—correctly—as easier to implement and—incorrectly—as more flexible than competitors. Next we discuss the GECM and review it under various scenarios. We show that common practices among political scientists lead to large problems with Type I errors for independent variables and a sometimes massive tendency towards claiming error correction.

3 The General ECM—Origins and Explanation

A single-equation general ECM was introduced to applied econometrics by Hendry and Anderson (1977) and Davidson et al. (1978), who each included level-form lags in differenced regressions to estimate equilibration.[8] Despite modeling individual $I(1)$ series, the relationship between them could take on a stationary equilibrium. With many economic and financial time series fitting the $I(0)/I(1)$ dichotomy—e.g., stock prices and exchange, interest, unemployment, and inflation rates—the model's assumptions were reasonable. Subsequent research by Engle and Granger (1987) formalized the concept as cointegration.

[7]Stock and Watson (2011, 650) study one-year and three-month interest rates as an example of cointegration. Some political series may be fractionally cointegrated (e.g., Lebo and Young 2009 and Box-Steffensmeier and Tomlinson 2000) but political data as close as Stock and Watson's are unlikely.
[8]Development of the method can be traced further back to Phillips (1954, 1957) and Sargan (1964).

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

In one form the bivariate ECM can be written as

$$\Delta Y_t = \beta_0 \Delta X_t + \alpha_1 (Y_{t-1} - X_{t-1}) + \epsilon_t \tag{3}$$

$$\Delta X_t = \upsilon_t, \tag{4}$$

in which ϵ_t and $\upsilon_t \sim N(0, \sigma^2)$.

In equation (3), current changes of ΔY_t can occur in response to ΔX_t or to correct dis-equilibrium between Y_{t-1} and X_{t-1}. With $I(1)$ series, testing whether $\alpha_1 = 0$ is equivalent to a cointegration test such that when $\alpha_1 < 0$ cointegration exists, i.e., the series are in a long-run equilibrium (Bannerjee et al. 1993). Further, with $I(1)$ data, ΔY_t and ΔX_t are stationary and, if cointegration exists, then $Y_{t-1} - X_{t-1}$ is as well.[9] Thus, with cointegration, equation (3) is *balanced*—a critical property for correct specification—and is safe from spurious regressions. Without cointegration, however, the model is unbalanced and the practitioner should set aside the estimates and choose a different specification.

Applying equation (3) is somewhat cumbersome, however, since post-estimation calculations are needed to obtain the most useful parameters, including the error-correction parameter. Consequently, Bårdsen (1989) derived the reparamaterization that is now the commonly used form:

$$\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 \Delta X_t + \beta_1 X_{t-1} + \epsilon_t. \tag{5}$$

In equation (5), both the ECM parameter (α_1) and the short-run effects of X_t are estimated directly while long-run multipliers are readily calculable. DeBoef and Keele (2008) highlight these features and much of the second half of *Taking Time Seriously* is devoted to exploring the potential of equation (5) in political science.[10] Long after its discussion in a 1992 special cointegration issue of *Political Analysis*, DeBoef and Keele's re-introduction drew significant attention to the GECM. Equation (5) has since become extremely popular—perhaps the most popular time-series model in political science.[11]

Yet, the GECM is consistently misused. Whereas D&K highlight the GECM when all series are stationary, researchers rarely pay close attention to the order of integration of their time series and often skip important tests. This has had serious inferential consequences. For example, political scientists have not recognized that if Y_t is $I(1)$, then the GECM is effectively a cointegration test. With Y_t as $I(1)$, *all* series in the model must be $I(1)$ and cointegration must be present. Otherwise, the model is unbalanced and misspecified.

Also missed is that the distribution of the ECM parameter's (α_1) test-statistic is neither standard-normal nor dimension invariant—it shifts systematically depending on sample size as well as the number and stationarity properties of the variables (Kremers, Ericsson, and Dolado 1992; Hansen 1995; Ericsson and MacKinnon 2002). The usual practice erroneously uses a standard one-tail t-test for hypothesis testing resulting in over-blown findings. Seemingly, equilibrium relationships have been found everywhere.

With strictly $I(1)$ data, one *can* calculate new critical values, however, and Ericsson and MacKinnon (2002) provide response surfaces for such calculations based upon the length of T and the number of covariates. Table 1 provides our calculations of critical values for lower values of T and up to five independent variables. Yet these values are only applicable if *every* series is strictly $I(1)$. Any diversion for *any* variable in the model alters the test statistic's sampling distribution and critical values—often dramatically. Additionally, any loss of equation balance

[9]Kremers et al. (1992) were the first to formally establish the relationship between the t-test on the disequilibrium term $Y_{t-1} - X_{t-1}$ and Engle & Granger's cointegration test-statistic; both are unit-root tests, but the GECM is a more powerful test for cointegration because it does not impose common factor restrictions. See Appendix A in Supplementary Materials for further background on the links between the GECM model, the Dickey–Fuller test, and the Engle and Granger two-step cointegration method.

[10]The equivalence of the ADL and the Bårdsen ECM is presented in DeBoef and Keele (2008, 189–91).

[11]In economics the method has been largely supplanted by alternatives like the ARDL-bounds tests of Pesaran, Shin, and Smith (PSS, 2001). As explained below, the PSS test allows flexibility for regressors' orders of integration but the dependent variable must still be $I(1)$. The PSS paper has been cited over 3000 times since publication in 2001 whereas Ericsson and MacKinnon (2002)—the source for the GECM's correct critical values—has been cited 243 times.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 1 The 5% MacKinnon critical values of ECM t-statistic for $I(1)$ Data

T	*1 IV*	*2 IVs*	*3 IVs*	*4 IVs*	*5 IVs*
35	−3.316	−3.613	−3.867	−4.082	−4.268
40	−3.300	−3.598	−3.85	−4.066	−4.255
45	−3.290	−3.587	−3.838	−4.055	−4.246
50	−3.283	−3.578	−3.829	−4.047	−4.239
55	−3.277	−3.570	−3.822	−4.040	−4.233
60	−3.270	−3.565	−3.816	−4.035	−4.229
65	−3.267	−3.560	−3.811	−4.030	−4.226
70	−3.263	−3.556	−3.807	−4.027	−4.223
75	−3.259	−3.552	−3.803	−4.023	−4.220
80	−3.256	−3.549	−3.800	−4.021	−4.218

Note. Critical values for each T computed using response surfaces provided by Ericsson and MacKinnon (2002). Estimates of critical values based on model with intercept and no trend.

makes a cointegration test dubious so, again, if the dependent variable is $I(1)$, then the model should only include $I(1)$ independent variables.

Yet, recent research on near- and fractional integration demonstrates that true unit-roots are rare in political data (Box-Steffensmeier and Smith 1996, 1998; DeBoef and Granato 1997; Byers, Davidson, and Peel 2000; Lebo, Walker, and Clarke 2000).[12] Thus, although Table 1's values allow proper testing of the ECM with all $I(1)$ series, the complications of political data mean that it is rarely applicable.[13]

Below, we examine the accuracy of inferences using common practices and the GECM. When α_1 is unknown and researchers estimate the model, how accurate are the conclusions? We study six types of dependent variables: (1) Y_t as a unit-root ($d = 1$ in a (p, d, q)) model, (2) Y_t as a bounded unit-root, (3) Y_t as stationary ($d = 0$), (4) Y_t as near-integrated, (5) Y_t as fractionally integrated, $(0 < d < 1)$, and (6) Y_t as explosive ($d > 1$).

For each type we simulate various covariate types—X_t as $I(1)$, $I(0)$, and fractionally integrated—and use 1–5 Xs with T = 60 and T = 150. In all, 746 simulation exercises test the GECM's abilities to estimate error correction and make other inferences. Absent the model being used as a cointegration test for purely $I(1)$ series with proper critical values, the GECM poorly measures the error correction parameter and finds it to be significant far too often—Type I errors occur at alarming rates. We begin with the case of Y_t as $I(1)$.

4 The General ECM Under Six Scenarios

4.1 *Case 1: The Dependent Variable Is a Unit-Root, I(1)*

What problems arise with the GECM and a unit-root Y_t? This scenario *can* work but there are four points that political scientists have missed: (1) X_t must also be a unit-root to ensure equation balance (Banerjee et al. 1993, 164–68), (2) Y_t and X_t must be cointegrated with each other (Banerjee, Dolado, and Mestre 1998; Ericsson and MacKinnon 2002), (3) non-standard critical values must be used (Ericsson and MacKinnon 2002) to evaluate cointegration, and (4) the ECM parameter is biased downwards as Xs are added to the model.

The value of the first two points is more easily seen in equation (3)'s version of the GECM: $\Delta Y_t = \beta_0 \Delta X_t + \alpha_1(Y_{t-1} - X_{t-1}) + \epsilon_t$. Here, if Y_t and X_t are unit-roots, then ΔY_t and ΔX_t are

[12] There are fields in political science where one might find more unit-root data, such as in international political economy, e.g., Bernhard and Leblang (2002), Leblang and Bernhard (2006).

[13] One caveat worth noting is that the model's estimates prove reliable if we specifically *generate* our data so that $\alpha_1 < 0$ and the ECM is used as a cointegration test. That is, in these specialized circumstances the ECM model does not make Type II errors, even with near-integrated data, and properly finds that cointegration is present (Kremers, Ericsson, and Dolado 1992; Hansen 1995; DeBoef and Granato 1999; Zivot 2000). But in actual practice, we do not choose the value of α_1 and, as Zivot (2000) shows with a single-equation ECM, α_1 is far better at *detecting* cointegration than it is at measuring the *strength* of re-equilibration (i.e., as an ECM).

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 2 Results of GECM model with $I(1)$ Data, $T=60$

Model	*1 IV*	*2 IVs*	*3 IVs*	*4 IVs*	*5 IVs*
ECM Significant – one tail t-distribution (%)	56.8	64.9	72.0	76.7	80.2
ECM Significant – MacKinnon values (%)	5.3	5.5	5.3	5.0	5.1
Mean of α_1	−0.12	−0.15	−0.18	−0.21	−0.24
Mean of α_1^*	−0.32	−0.37	−0.42	−0.47	−0.52
ECM & $\geq 1\Delta X_t$ Significant (%)	3.8	8.6	13.3	18.4	24.3
ECM & $\geq 1X_{t-1}$ Significant (%)	14.1	25.0	34.2	41.1	47.1

Note. Percentage results in each cell based on 10,000 simulations per ECM model with null hypothesis of no error correction. Finding of significant ECM and significant X_{t-1} indicates presence of error correction.
Mean of α_1^* when t-statistic exceeds MacKinnon critical value, i.e., cointegration is present.
ΔX_t and X_{t-1} significance ($p \leq 0.05$, two-tail test). ECM significance ($p \leq 0.05$, one-tail test).
Dependent Variable (DV) and all Independent Variables (IVs) are unit-roots ($I(1)$). $T = 60$.
ECM Model: $\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 \Delta X_t + \beta_1 X_{t-1} + \epsilon_t$.

stationary. And, if they are cointegrated, then $(Y_{t-1} - X_{t-1})$ is stationary as well.[14] Ordinary Least Squares (OLS) can be used here but it has been shown that the hypothesis test on α_1 in equation (5) is non-standard (Ericsson and MacKinnon 2002). The correct values can be found in common sources like Enders (2010, 493) or calculated for specific lengths of T as we do in Table 1.

Unfortunately, practitioners are neither paying proper attention to their series' stationarity nor testing for cointegration.[15] In particular, one often-quoted point in *Taking Time Seriously* seems universally misinterpreted: "Alternately, as the ECM is useful for stationary and integrated data alike, analysts need not enter debates about unit-roots and cointegration to discuss long-run equilibria and rates of reequilibration" (p. 199). This statement must be carefully interpreted: the GECM can work for *all* stationary *or all* integrated data. If all data are stationary, then one does not have to test for cointegration to estimate the GECM. But researchers have assumed this means that they need *never* worry about stationarity or cointegration—that plugging a mix of data into the GECM is fine and that the model can always work with or without cointegration.[16] But if the right-hand-side variables are not truly $I(1)$, the equation is not balanced and, even if balanced, if the series are not truly cointegrated, then problems arise.

In the absence of cointegration, having unit-roots Y_{t-1} and X_{t-1} on the right-hand side causes problems for hypothesis testing their coefficients. In common practice, a significant α_1 is used to confirm cointegration and its value is given as the speed with which a long-run equilibrium between Y_t and X_t is restored following a shock that separates them. Since α_1 is also used to calculate long-run multipliers for the covariates, bias and/or a Type I error gives a lot of false information.[17]

Table 2 shows GECM simulations using all $I(1)$ data (T = 60). *All* political science applications of the GECM have used the t-distribution for α_1's hypothesis test. The first row shows Type I error rates when this mistake is made.[18] With a single $I(1)$ X, 56.8% of ECM coefficients are significant at the

[14]D&K allude to this in their Footnote 11 guideline: "When using an ECM with integrated data, analysts must ensure that all terms on the right-hand side of the equation are stationary" (p. 190).

[15]The Granger Representation Theorem (Engle and Granger 1987) posits that if cointegration exists, then Granger causality must be present in at least one direction. Failing to explicitly test for cointegration or Granger causality can lead to erroneous causal claims.

[16]For example, Volscho and Kelly (2012) say: "...ECMs accommodate stationary and integrated variables, which is useful because our analysis has a mix of both data types. In summary, the ECM is a very general model that is easy to implement and estimate, does not impose assumptions about cointegration, and can be applied to both stationary and nonstationary data (Banerjee et al. 1993; De Boef and Granato 1999; De Boef and Keele 2008)." See Appendix F in the Supplementary Materials for other misinterpretations.

[17]For an example of this in action, see Table D.9 in the Supplementary Materials.

[18]Simulations were run in RATS 8.0. Series were specified to have no relationship to each other. $I(1)$ series were generated as $Y_t = Y_{t-1} + \epsilon_t$. $I(0)$ series were generated as $Y_t = \epsilon_t$. $I(d)$ series were generated to a specified order of d using the RATS "arfsim" package. Near-integrated series were generated by specifying $\rho \sim (0.9 - 1.0)$ in $Y_t = \rho Y_{t-1} + \epsilon_t$. Bounded series were created following Nicolau (2002). A brief explanation of related versus unrelated series can be found in Appendix B.1. Details on procedures for bounded series are in Appendix B.2 of the Supplementary Materials. Complete tables of results are in Appendix G.1 (near-integration, one table each for $T = 60$ with 1,2,3,4, and 5 IVs and one table each for $T = 150$ with 1,2,3,4, and 5 IVs) and Appendix H.3 (fractional integration, one table each for $T = 60$ with 1,2,3,4, and 5 IVs and one table each for $T = 150$ with 1,2,3,4, and 5 IVs).

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

0.05 level using a one-tailed t-test.[19] With correctly calculated critical values for α_1, Type I errors on both the ECM parameter and the Xs can be minimized. The second row shows just this—correct MacKinnon values put Type I errors where they should be for α_1, around 5% with 1–5 covariates.

But using incorrect critical values makes one likely to falsely conclude that cointegration exists and that the equation is balanced. With X_{t-1} as $I(1)$, unresolved auto-correlation leads to bias in the estimation of β_1, shrunken standard errors, and higher Type I error rates—β_1 is significant (0.05 level) 14.1% of the time in a bivariate model. Additional $I(1)$ Xs increase these problems and Type I errors creep higher. The row for ΔX shows that even though appropriately filtered, auto-correlation elsewhere gives us Type I errors on β_0 far too often. A model can begin to look quite good with both a significant ECM parameter and a long-run X. Using a standard distribution can lead one there far too easily.

Another issue is that additional $I(1)$ Xs move the distribution of α_1 further from zero. Even when the proper MacKinnon values indicate the presence of cointegration, the expected value of α_1 decreases with each additional covariate. Ease of interpretability is a key selling point of the GECM but this is one of several complications ignored by practitioners.

In all, Table 2 shows problems with the GECM with unrelated unit-root series. Testing for cointegration and relying on MacKinnon values can address spuriousness but we might still misstate the rate of error correction in a multivariate model. And, in practice, this ideal scenario is exceedingly rare—not only would the data need to be truly $I(1)$ (rare enough) and actually cointegrated (rarer still), they would also need to be unbounded.

4.2 *Case 2: The Dependent Variable Is a Bounded Unit-Root*

Time series are often *bounded* (sometimes called "limited" or "regulated") between an upper and lower limit.[20] Approval of leaders and parties, indices of economic evaluations like the ICS, Policy Mood, and percentages of political and economic phenomenon all fluctuate between upper and lower limits. Little is known about the consequences of bounded series—especially regarding their effects on error correction models.

A bounded time series is an odd case. For one, it cannot meet the textbook definition of a unit-root since its asymptotic properties include mean reversion and finite variance (Williams 1992). Series like presidential approval and Stimson's *Mood* (1991) cannot have infinite variance or break their bounds and will oscillate around a long-term mean. Still, over long periods a bounded series can exhibit the perfect-memory of integrated data or the long-memory of near- and fractionally integrated data. Indeed, recent work has shown that boundedness is compatible with unit-root properties and the idea of a *bounded unit-root*, $BI(1)$, exists in econometrics (Cavaliere 2005; Granger 2010).

Problems occur when the asymptotic properties of a bounded series are used to dismiss the possibility of a unit-root and treat a series as stationary.[21] Bounded series also over-reject the Dickey–Fuller test (1979), increasing the tendency to treat them as stationary (Cavaliere and Xu 2014). If one treats stationarity as a yes/no question and models a bounded series as simply stationary, then auto-correlation remains and inferences can be threatened.

Here we are interested in the problems bounded series present to the GECM. Bias comes about simply: if Y_{t-1} was near the series' bounds, ΔY_t will have a strong tendency in the opposite direction. High (low) levels at Y_{t-1} will be followed by negative (positive) changes at t, pushing α_1 further into negative territory. This may tell us something about Y_t but it misinforms our inferences

[19] Of the GECM articles we surveyed, all used a standard t-test and averaged roughly four IVs per regression: Faricy (2011): 4, Jennings and John (2009): 2 plus interactions, Kayser (2009): 6–7 plus interactions (with a larger N), Kelly and Enns (2010): 3–4, Kono (2008): 9 (with a very large N), Ramirez (2009): 9, Rickard (2012): 5, Ura (2014): 4, Ura and Wohlfarth (2010): 4, and Ura and Ellis (2012): 5.

[20] Series bounded on one end are also problematic (Cavaliere 2005; Granger 2010).

[21] Durr (1992) pushes analysts to focus on the data in hand and rely on theory for whether series might be integrated or not. Researchers who are unaware or dismissive of fractional integration techniques and who also side with Williams's (1992) argument will assume bounded series like Congressional Approval cannot be $I(1)$ and will routinely treat them as $I(0)$. Keele (2007, 252) is pragmatic on this point and tests for stationarity and fractional integration while noting that asymptotically the series cannot be integrated.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 3 GECM estimation problems with a bounded dependent variable

	Bivariate					
DV bounds	(1 – 100)	(49 – 71)	(1 – 100)	(49 – 71)	(1 – 100)	(49 – 71)
model type	$T=60,\ \sigma=1$		$T=60,\ \sigma=2$		$T=60,\ \sigma=3$	
ECM significant %	60.0	65.3	61.4	75.3	62.6	86.6
ECM significant[a] %	6.0	7.7	6.1	9.9	6.1	12.4
Mean of α_1	−0.12	−0.14	−0.13	−0.17	−0.13	−0.20
	$T=100,\ \sigma=1$		$T=100,\ \sigma=2$		$T=100,\ \sigma=3$	
ECM significant %	61.2	70.0	63.1	87.0	64.8	98.3
ECM significant[a] %	7.0	10.5	7.6	14.8	8.1	23.1
Mean of α_1	−0.08	−0.10	−0.08	−0.13	−0.09	−0.16
	$T=150,\ \sigma=1$		$T=150,\ \sigma=2$		$T=150,\ \sigma=3$	
ECM significant %	61.5	74.3	64.3	97.3	67.0	99.9
ECM significant[a] %	6.5	12.2	7.6	20.5	8.4	45.6
Mean of α_1	−0.05	−0.07	−0.06	−0.10	−0.06	−0.14
	Multivariate (2 IVs)					
DV bounds	(1–100)	(49–71)	(1–100)	(49–71)	(1–100)	(49–71)
model type	$T=60,\ \sigma=1$		$T=60,\ \sigma=2$		$T=60,\ \sigma=3$	
ECM significant %	67.7	73.3	68.8	82.4	70.0	90.1
ECM significant[a] %	6.2	7.8	6.4	10.6	6.6	12.6
Mean of α_1	−0.16	−0.18	−0.16	−0.21	−0.17	−0.24
	$T=100,\ \sigma=1$		$T=100,\ \sigma=2$		$T=100,\ \sigma=3$	
ECM significant %	69.7	78.0	71.6	91.0	73.3	98.6
ECM significant[a] %	6.2	9.6	6.6	13.9	7.0	20.8
Mean of α_1	−0.10	−0.12	−0.10	−0.15	−0.11	−0.18
	$T=150,\ \sigma=1$		$T=150,\ \sigma=2$		$T=150,\ \sigma=3$	
ECM significant %	69.5	80.5	72.0	97.6	74.4	100
ECM significant[a] %	6.8	11.5	7.7	18.5	8.4	38.5
Mean of α_1	−0.07	−0.09	−0.07	−0.12	−0.07	−0.15

Note. Percentage results are of 10,000 simulations for each specified model and T. All independent variables are integrated $I(1)$. [a]Significant using MacKinnon values.

of re-equilibration between Y_t and X_t. Movement away from the bounds and towards the long-term mean will move α_1 but this is bias, *not* error correction.

This is shown in Table 3. Following Nicolau (2002), we simulate unit-roots that bounce back as they near upper or lower thresholds.[22] We explored the effects of tighter bounds, longer series, series with more variance, and models with additional independent variables. In all, $BI(1)$ series are problematic for estimating the ECM coefficient: the average value jumps and Type I errors rise. Also, the more the bounds come into play—tighter bounds, more variance, or a longer sample—the more Type I errors increase.

Next, we create series meant to mimic derivatives of Stimson's Mood (e.g., Kelly and Enns, 2010)—$BI(1)$ with $\sigma=3$ and bounds of 49 and 71—and regress them on an $I(1)$ X in the GECM. With $T=150$ the ECM parameter is significant 99.9% of the time using a one-tail t-distribution and 45.6% of the time with MacKinnon values.[23] Boundedness does not seem to affect estimation of β_1 or

[22]The process is generated by: $X_t = X_{t-1} + e^k(e^{-\alpha_1(X_{t-1}-\tau)} - e^{\alpha_2(X_{t-1}-\tau)}) + \epsilon_t$ where $\alpha_1 \geq 0$; $\alpha_2 \geq 0$; $k<0$; and ϵ_t is assumed to be *i.i.d.* with a mean of 0 and variance of 1. Appendix B.2 in the Supplementary Materials provides additional detail and graphics on how these series were generated.

[23]Transformation of a bounded dependent variable does not seem to solve the Type I error problem. We transformed our series into log odds as $Y_t^* = log\frac{Y_t}{1-Y_t}$, which creates a series with an unbounded error process. But the Type I error rate actually worsens—with $T=150$, bounds of 49 to 71, and $\sigma=3$, Type I errors with MacKinnon values increase approximately 2%.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 4 ECM significance and coefficient size by degree of DV autoregression

ρ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
ECM Significant (%)	100	100	100	100	100	100	100	99.8	97.7	79.1
Mean of α_1	−1.03	−0.92	−0.83	−0.73	−0.64	−0.54	−0.45	−0.36	−0.26	−0.17

Note. Cell entries are the result of 10,000 simulations for each bivariate model with the DV specified at ρ. IV integrated $I(0)$. ECM significance ($p \leq 0.05$, one-tail test).

β_0 but, if we use α_1 to discuss equilibrium relationships or to calculate long-run multipliers, we will misstate the relationships between the variables. Note that Table 3 represents the GECM's *best* chance with a bounded Y. A bounded and fractionally integrated Dependent Variable (DV) will be even more problematic than the $BI(1)$ series tested here. And, if X were to deviate from $I(1)$, further problems would occur.

This adds an important caveat to the GECM model. Even if we find series that are strictly unit-roots and we use MacKinnon CVs, mistakes are still rampant if our dependent variable is one of the vast majority of political times series that is bounded. Just how prevalent are bounded series? For instance, we chose our five paper replications based on their journal prominence but, as it happens, the papers use a total of 13 series as dependent variables and *all* 13 are bounded.

4.3 *Case 3: The Dependent Variable and All Independent Variables Are Stationary*

Next we review the case where all series are stationary, precisely the type of data DeBoef and Keele (2008) had in mind; their point that researchers may investigate long-run dynamics without worrying about cointegration is meant to apply to stationary data. Further, the use of a standard t-test for the error correction hypothesis is based on the assumption of stationarity. However, without adjusting how the GECM is interpreted with stationary data, a number of problems remain.

First, a stationary dependent variable can usually guarantee a significant ECM parameter. Second, in many instances, neither the size of the ECM coefficient nor the strength of its significance imply a close relationship between the dependent variable and other series in the model. As a demonstration, Table 4 shows the simulated results of a bivariate GECM in which the DV is stationary but with varying degrees of auto-regression, i.e., (ρ,0,0), and the Independent Variable (IV) is a white noise process, i.e., (0,0,0).

The prevalence of significant ECMs indicates that we should rethink the meaning of the α_1 coefficient when data are stationary. We are seeing re-equilibration here, but it is the natural reversion of the dependent variable to its mean. That is, α_1 is capturing the speed with which Y returns to *its own* equilibrium (its mean) following a shock or an effect from X. What we see in Table 4 is not evidence of error correction—it is simply confirmation that Y is stationary.[24] When working with stationary data, the larger the absolute value of the ECM, the *less* of a long-run relationship between X and Y. If Y has no long-run tendencies except mean reversion, it is an odd choice to be looking for its long-run relationship with X.

This is problematic when common practice is to use α_1 to speak to error correction and re-equilibration *between* variables. This occurs frequently in practical applications—a statistically significant and substantively large α_1 coefficient is claimed to indicate strong re-equilibrating behavior between Y_t and some independent variables with weaker connections.[25] Using the raw

[24]This is not surprising given that all stationary series are in equilibrium, as recognized by (DeBoef and Keele 2008, 189) and (Banerjee et al. 1993, 4).

[25]Casillas, Enns, and Wohlfarth (2011) investigate public opinion's effect on salient Supreme Court decisions. Finding an ECM of −1.27 (s.e. = 0.15), they say: "The magnitude of the error correction rate in this model suggests that, following just one term, the Court's behavior almost completely adjusts to changes in ideology and social forces at term t." And, (footnoted): "Because the error correction rate indicates the proportion of the long-term effect that occurs in each subsequent time period, an absolute value greater than 1 seems surprising." We find *Salient Cases* to be stationary (d = 0.3 (s.e. = 0.08); DF = 3.98*) in level form, which explains how the ECM value is possible. A second example, Jennings and John (2009), uses agenda items in Queen's Speeches 1960–2001—close to stationary—as dependent variables and reports many ECMs below −1.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 5 The ADL versus the GECM: Effects of congressional approval on Supreme Court approval

Model	*ADL*	*GECM*
Supreme Court approval$_{t-1}$	0.27 (0.16)	
Error correction$_{(SupremeCourtApproval_{t-1})}$		−0.73* (0.16)
Congressional approval	0.49* (0.22)	
Congressional approval$_{t-1}$	−0.05 (0.24)	0.43* (0.15)
Δ Congressional approval		0.49* (0.22)
Constant	41.83* (12.48)	41.83* (12.48)
N	41	41
R^2	0.41	0.43
Adjusted R^2	0.36	0.39

Note. Data comes from Durr, Martin, and Wolbrecht (2000). ADL Model: DV *Supreme Court Approval* represents the authors' semi-annual measure of approval for the Court. ECM Model: DV Δ *Supreme Court Approval* represents changes to this measure of approval. Significance of ECM (*p < .05, one-tail test) and coeffcients (*p < .05, two-tail test).

ECM value to compute long-run multipliers would further solidify a researcher's faith in the (falsely strong) relationship.

Problems extend beyond mistaken inferences regarding α_1. To explain, we return to the linear transformation between the ADL and GECM. Despite the algebraic equivalence of the two models, as D&K note, "differing quantities are directly estimated in each model" (DeBoef and Keele 2008, 195). The autoregressive distributed lag (ADL) model is specified as

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t, \tag{6}$$

and the Bårdsen GECM transforms this into

$$\Delta Y_t = \alpha_0 + \alpha_1^* Y_{t-1} + \beta_0^* \Delta X_t + \beta_1^* X_{t-1} + \epsilon_t. \tag{7}$$

D&K (p. 190) note the equivalence between parameters in the ADL and the GECM, specificially: $\alpha_1^* = (\alpha_1 - 1)$, $\beta_0^* = \beta_0$ and $\beta_1^* = \beta_0 + \beta_1$. Here we are interested in the long-run parameter, β_1^*. It is the sum of two quantities from the ADL ($\beta_0 + \beta_1$) but, critically, its standard error is not.

For example, we estimate bivariate ADL and GECM models using data from Durr, Martin, and Wolbrecht (2000): the DV is *Supreme Court Support* and the IV is *Congressional Approval*, both of which are found to be stationary by the authors (fn. 13, p. 772). Our results are in Table 5.

Equation (6)'s ADL has two dynamic parameters, α_1 and β_1. The short-term effect is captured as β_0 and the long-run effect can be calculated as $(\beta_0 + \beta_1)/(1 - \alpha_1)$. What we see in the first column of Table 5 is that the dynamic parameters fail to reject their null hypotheses—neither the lag of the dependent variable ($\alpha_1 = 0.27$, s.e. = 0.16) nor the lagged IV ($\beta_1 = -0.05$, s.e. = 0.24) are significant. In short, there is no support for the use of a dynamic model.

But looking at the GECM, we see that the error correction parameter—α_1^* calculated as the ADL's α_1 minus 1—*is* significant, as we would expect from Table 4. Additionally, because β_1^* consists of the ADL's $\beta_0 + \beta_1$ while the standard error is not additive, the long-run effect of *Congressional Approval* is now significant at the $p < 0.05$ level. Whereas the results of the ADL conclude that we should estimate a static regression, the significant parameters in the GECM encourage the researcher in the opposite direction. Given publication bias towards significant findings, the GECM is again attractive. The ADL and the GECM may be algebraically equivalent, but the reorganization of parameters is not benign.

4.4 *Case 4: The Dependent Variable Is Strongly Autoregressive / Near-Intergrated*

D&K argue that "the only situation where one would strongly prefer the ECM [as opposed to the ADL] is if the data are strongly autoregressive" and "[because the] variables are parameterized in terms of changes, helping us to avoid spurious findings if the stationarity of the series is in question

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 6 Rejection rates for the null hypothesis of ADL ($\beta_1 = 0$) and GECM ($\alpha_1^* = 0$ & $\beta_1^* = 0$) with near-integrated data

	ρ_x					
ADL model						
ρ_y	0.75	0.80	0.85	0.90	0.95	0.99
0.75	5.7	5.8	5.6	5.6	5.9	6.0
0.80	5.7	5.9	5.7	5.7	5.9	6.0
0.85	5.7	5.9	5.8	5.8	6.0	6.0
0.90	5.7	5.9	5.8	5.8	6.0	6.1
0.95	5.7	5.9	6.0	5.8	6.1	6.3
0.99	6.0	5.9	6.0	5.8	5.9	6.2
GECM model						
ρ_y	0.75	0.80	0.85	0.90	0.95	0.99
0.75	7.6	8.1	8.3	8.8	9.1	9.3
0.80	7.9	8.4	8.9	9.4	9.9	10.2
0.85	8.2	8.8	9.2	10.2	10.7	11.4
0.90	8.1	8.7	9.6	10.8	11.8	13.0
0.95	7.3	8.0	9.2	10.6	12.3	14.3
0.99	6.4	7.1	8.1	9.7	11.4	13.7

Note. Percentage results based on 10,000 simulations. $T = 60$. Each model contains one IV.
ADL model: $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$
GECM: $\Delta Y_t = \alpha_0 + \alpha_1^* Y_{t-1} + \beta_0^* \Delta X_t + \beta_1^* X_{t-1} + \epsilon_t$. Significance of α_1^* (*$p < .05$, one-tail test)

due to strongly autoregressive or near-integrated data" (DeBoef and Keele 2008, 195). We find, however, that the GECM runs into serious problems here as well. In particular, rejection rates on the long-run parameter (β_1^*) reach unacceptable levels with strongly autoregressive, or near-integrated series.

DeBoef (2001) and DeBoef and Granato (1999) discuss the GECM in the case of near-integrated data and find it appropriate. Additionally, DeBoef and Granato (1997) investigate rates of spurious regressions with near-integrated data and find that the ADL solves the spurious regression problem. However, this is not the case for the GECM—spurious regressions appear too frequently with either near-integrated or other stationary data. With the DGP of $Y_t = \rho Y_{t-1} + \epsilon_t$, we find elevated rates of Type I errors on β_1^* far below the range of ρ that is typically considered near-integrated, i.e., 0.90–0.99.

Our findings are different from the previously mentioned studies for a number of reasons. First, the investigation of the power of the GECM in both DeBoef (2001) and DeBoef and Granato (1999) uses models in which the data are pre-specified to be (near-)cointegrated. As we note above, the GECM is far more powerful at detecting cointegration (avoiding Type II errors) than it is at estimating its presence (avoiding Type I errors) (Zivot 2000). Second, as noted, the ADL and GECM do not estimate the same parameters, and as a result, they produce different rejection rates. Whereas the ADL is suitable for near-integrated series, fitting the exact same data in a GECM produces spurious regressions. These problems are persistent, even with dependent variables that are clearly stationary.

To investigate the disparity in rejection rates between the ADL and the GECM, we generate series with varying degrees of autoregression. The upper section of Table 6 contains the rejection rates of the null hypothesis that β_1 of the ADL model is equal to zero.[26] The lower section contains

[26]This is a recreation of Table 4 of DeBoef and Granato (1997), which tested the rejection rate of β_0. In our case, we are interested in the rejection rates of β_1, the coefficient on X_{t-1} in each model. Additionally, we have expanded our criteria to include series with lesser degrees of autoregression. Neither of these changes affect the ultimate conclusion with respect to the ADL: the model is proper for use with near-integrated data.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

the rejection rates for the null hypothesis with respect to the GECM: that both the error correction parameter α_1^* and the long-run parameter β_1^* of the GECM model are jointly equal to zero.[27]

Our findings for the ADL match those of DeBoef and Granato (1997), who find that the model has acceptable spurious regression rates with near-integrated data. But we also find that this does not translate for the same data in the GECM. Precisely the type of data D&K claim we should favor with the GECM proves susceptible to Type I errors. Excessive rejection rates are also found when data are less strongly autoregressive, e.g., $\rho = .75$.[28]

There are a number of important conclusions to draw here. First, despite being algebraically equivalent, the GECM and the ADL do not produce the same results. The GECM does not perform as well with strongly autoregressive, or near-integrated data. In terms of model preference, the ADL should be preferred over the GECM with either stationary, strongly autoregressive, or near-integrated data. With a stationary DV the error correction parameter is practically guaranteed significance and the substantive meaning of its coefficient is inscrutable.

Second, although using MacKinnon CVs with near-integrated data would limit the rate of spurious regressions (see Section G.1 of Supplement), this cannot be recommended since the decision of when to switch to MacKinnon values with stationary data will be an arbitrary one. The MacKinnon values are recommended based on the unit-root or not distinction. Researchers cannot simultaneously argue that data are stationary while using unit-root critical values. Spurious regressions appear, for example, when $\rho = 0.75$ and the correct critical values in that case are derived from neither the MacKinnon nor the normal distribution—they are unique to the particular data being used. Of course, series that are not unit-roots may not be simply autoregressive. Many political time series have been shown to be fractionally integrated and these prove even more troublesome for the GECM.

4.5 *Case 5: The Dependent Variable Is Fractionally Integrated, (0,d,0), and 0 < d < 1*

Fractional integration (FI) is common among political time series (Box-Steffensmeier and Smith 1996, 1998; Byers, Davidson, and Peel 2000; Lebo, Walker, and Clarke 2000; Lebo and Moore 2003), so finding problems here should make political researchers especially wary of the GECM.[29] Figure 2 shows a FI series and its ACF and PACF. Auto-correlations persist for long periods and, without pre-whitening, cause problems in regression analyses (Lebo, Walker, and Clarke 2000; Tsay and Chung 2000). What are the consequences of using fractionally integrated series in the GECM?

Recall that the GECM model is $\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 \Delta X_t + \beta_1 X_{t-1} + \epsilon_t$. We showed that when Y_t is a unit-root, the left-hand side of the equation is properly treated and ΔY_t is clear of auto-correlation. We also showed that if Y_t begins as stationary, applying the differencing operator Δ will *over*-difference it. In the FI case, with $0 < d < 1$, the GECM will again over-difference the series to varying degrees depending on the initial value of d. For example, a series that in levels is $d = 0.7$ will become $d = -0.3$.[30]

Over-differencing builds a moving average process into the transformed series (Dickinson and Lebo 2007). For example, if we (1) begin with a FI series: $(1-L)^d Y_t = \epsilon_t$, where $\epsilon_t \sim N(0, \sigma^2)$, (2) isolate Y_t, and (3) apply first differencing, we get: $\Delta Y_t = \frac{\epsilon_t}{(1-L)^d} - \frac{(\epsilon_{t-1})}{(1-L)^d}$. If $d = 0$, this leaves a non-invertible moving average process as discussed above. With smaller degrees of over-differencing, an

[27]To reject the null hypothesis, both α_1^* and β_1^* must be significant.

[28]Since these results are of a joint hypothesis test, two parameters must be significant. The rejection rates of just the β_1^* are higher with near-integrated data. Additionally, while these results are based on a bivariate model, in practice the GECM is fit with more IVs. The probability of finding a significant β_1^* parameter increases with the inclusion of additional IVs. Further, the nature of the ECM with stationary data means the significance of the α_1^* parameter is not a sufficient gatekeeper.

[29]Aggregating individuals with heterogeneous AR processes creates a long-memory series with a mix of stationary and non-stationary attributes (Granger 1980). Many political time series are created in this way. Pickup (2009) is the only attempt we know of that argues that FI does not apply to aggregate political data (but see Young and Lebo [2009] for a rebuttal).

[30]DeBoef and Granato (1997) discuss the near-integrated case and find that an over-differenced series will contain negative auto-correlation in its error term and will produce non-credible inferences.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

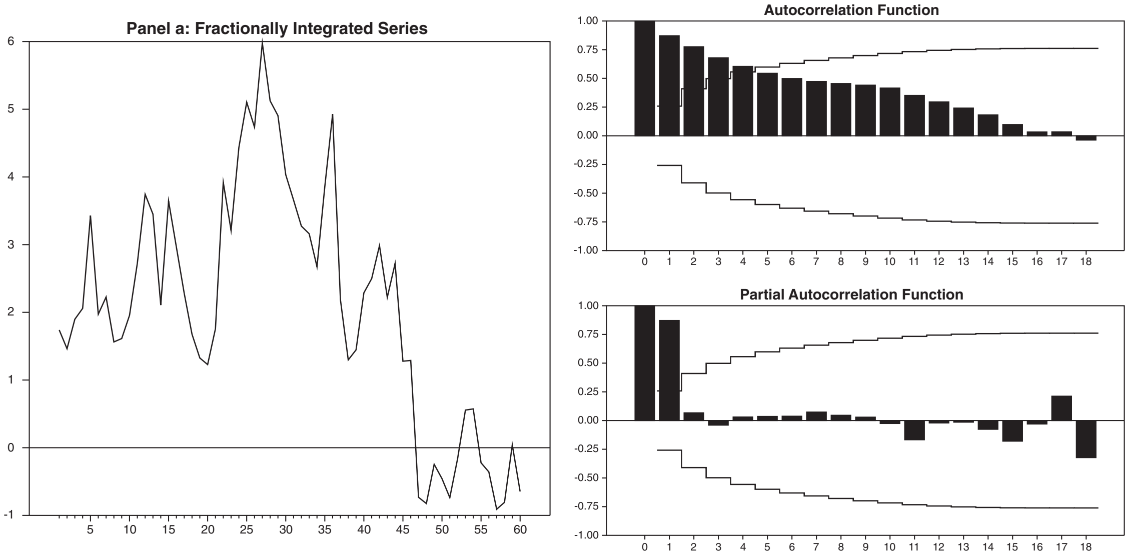

Fig. 2 A fractionally integrated series with ACF and PACF.
Note. Significant spikes on the ACF and PACF are problematic correlations.

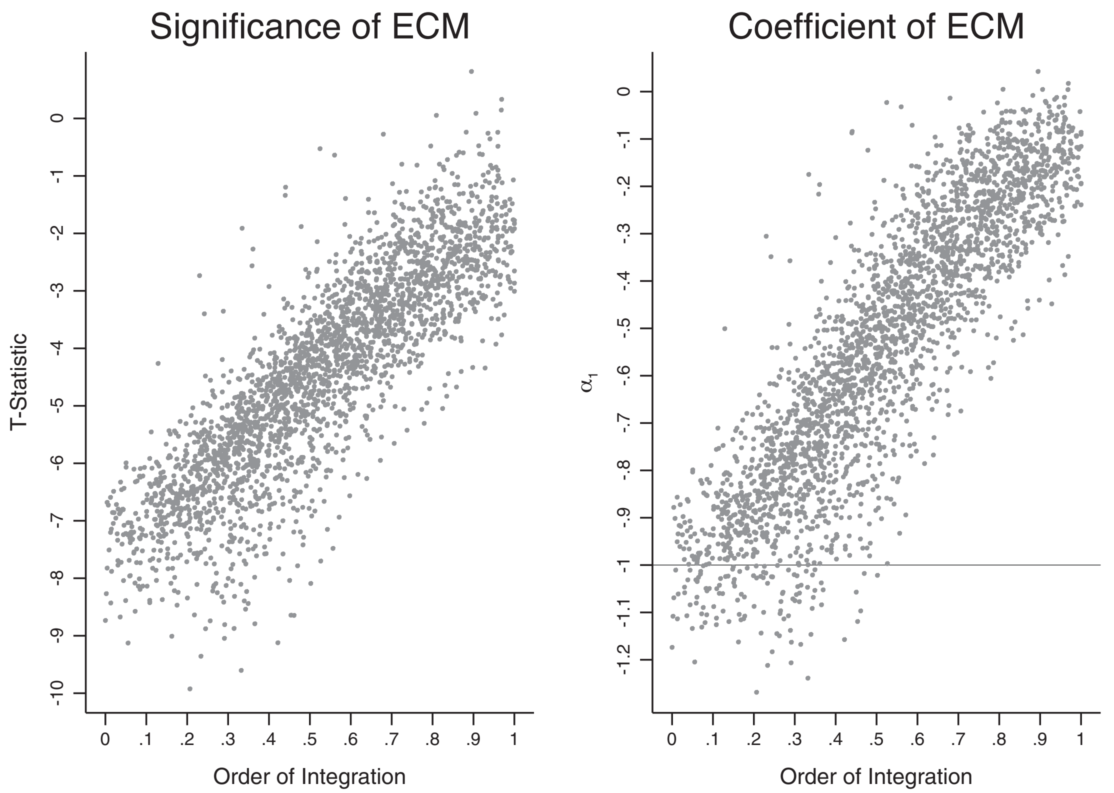

Fig. 3 Significance and size of ECM parameter based on Y's order of fractional integration.
Note. Plots based on bivariate regressions with $T = 60$. Horizontal line on the right indicates the theoretical limit of an ECM.

MA process of $\frac{1}{(1-L)^d}$ becomes a part of ΔY_t. Over-differencing by just 0.4, for example, gives a 99.4% chance of creating a significant MA parameter (Dickinson and Lebo 2007).

Figure 3 shows the consequences of first-differencing FI data in the GECM. The X-axis shows Y_t's original order of integration. On the left panel's Y-axis is the ECM's t-statistic and on the right

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

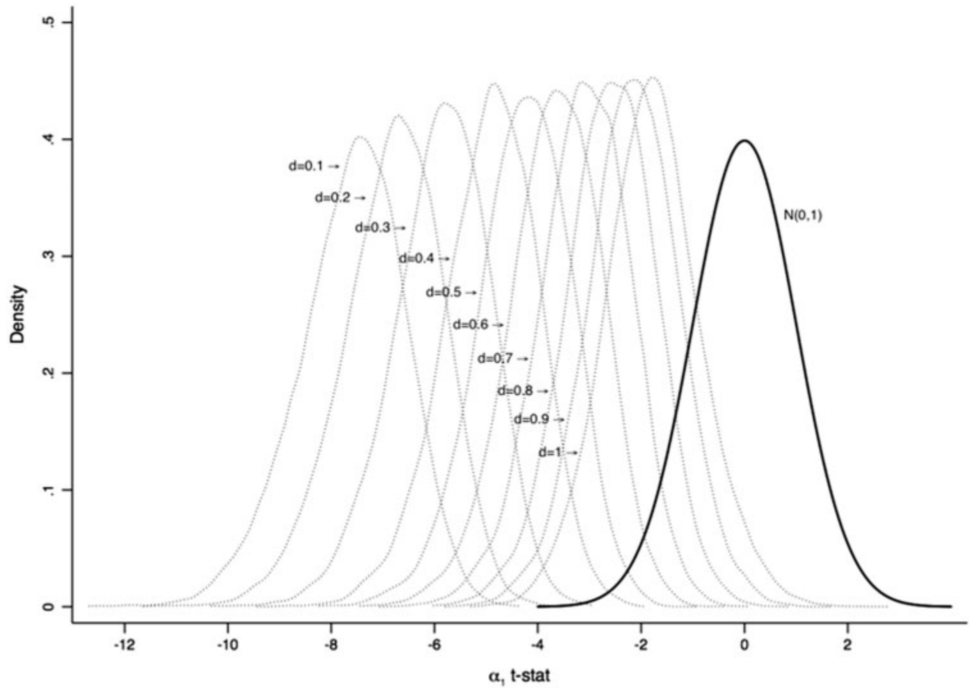

Fig. 4 Sampling distribution of the ECM *t*-statistic by order of fractional integration.
Note. Density plots of *t*-distributions from simulations of bivariate GECMs when both DV and IV are FI, *I*(*d*). $T=60$. ECM Model: $\Delta^d Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 \Delta^d X_t + \beta_1 X_{t-1} + \epsilon_t$.

panel's Y-axis is the ECM coefficient.[31] By chance, the probability of getting a *t*-statistic of -4 or less is roughly 1 in 10,000 and -7 or less is roughly 1 in 100,000,000. But such *t* values are routine when using unrelated FI data in the GECM. As Y gets further into stationary territory, the coefficient for Y_{t-1} declines and Type I errors become the norm.

The right-hand panel of Fig. 3 shows the size of the ECM coefficient. Read as the (negative of the) proportion of the gap that is re-equilibrated in the next period, it should never be below -1. Here we see that it often is—a fact that might lead a researcher to conclude they have found an especially close relationship between Y_t and the Xs.

For a different view, Fig. 4 shows the density plots of the ECM parameter's *t*-statistic as Y_t's d value decreases from one. The sampling distribution for the test-statistic moves along with the order of integration of Y_t. Even if we could pin down the correct critical values, the meaning of the ECM coefficient has been lost. We are basically sliding toward the ECM story told above for a Y_t that is stationary. Ultimately, the value of α_1 tells us more about the level of memory in Y_t than about its relationship to independent variables.

4.6 *Case 6: The Dependent Variable Is Explosive, $d > 1$*

A time series where $d > 1$ is referred to as an explosive process so that increases (or decreases) are followed by larger increases (or decreases) (Fuller 2009). While not a common trait, explosive processes can be found in political science.[32] For our purposes here, the explosive case presents its own set of problems to the GECM, the most basic of which is that the equation will not be

[31] Here $T=60$ and X is $I(1)$. See Appendix H.3 and Tables H.3–H.12 in the Supplementary Materials for detailed results where we vary sample size and the order of integration for both X and Y. See Figure H.1 in the Supplementary Materials for another view of bias in the ECM in this case.

[32] For example, our estimate of d for the *Policy Liberalism* variable of Kelly and Enns (2010) is 1.35 (s.e. = 0.10) (see Table E.15 of Appendix E.2). This might indicate that the ideological tone of legislation passed by Congress shifts quickly following electoral gains and losses.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

balanced if one variable is explosive. When $d > 1$, using first-differencing for Y will leave it under-differenced and some measure of auto-correlation will remain. Additionally, having Y_{t-1} on the right-hand side will introduce a great deal of auto-correlation to the model. In short, explosive data are in need of pre-whitening and, when possible, some fractional filter allowing $d > 1$ will be the best approach to remove auto-correlation and allow reliable inferences (Hosking 1981).

4.7 *What Models Should We Use Instead?*

Our findings indicate that if one can establish that all series in the model are stationary, the ADL model is preferable to the GECM. If all series contain unit-roots, then one can look for cointegration and, if it is present, estimate a GECM while thinking about the consequences of bounds. Next, we discuss two more flexible ECM frameworks: the fractional cointegration and 3-step fractional ECM model of Clarke and Lebo (2003) and the auto-regressive distributive lag (ARDL) bounds approach of Pesaran, Shin, and Smith (2001, PSS).[33]

While the Engle and Granger (1987) approach begins with two $I(1)$ series combining to create a series of $I(0)$ residuals, fractional cointegration adds flexibility by relaxing the assumptions that (a) stationarity should be dealt with as a dichotomy, (b) the parent series need be $I(1)$, and (c) the residuals between Y and X need be $I(0)$ (Cheung and Lai 1993; Baillie and Bollerslev 1994; Dueker and Startz 1998; Box-Steffensmeier and Tomlinson 2000). To find fractional cointegration, the researcher must first establish that the parent series are of the same order of fractional integration, and second, that the ECM is of a lower order of integration (d) than the parent series.

The first step follows Engle and Granger (1987) and regresses the level-form Y on the level form X (or Xs) hypothesized to be error correcting with Y. The residual series is the ECM. The d value is then identified for each series—Y, X, and the ECM—using an estimator such as that of Robinson (1995).[34] If the d value for the ECM is less than the d value for *both* Y and X, then one can conclude that error correction is occurring. As with Engle and Granger (1987), not finding any evidence of error correction should be enough evidence to drop an ECM specification at this point. However, if one wishes, the approach is capable of following D&K's advice and estimating error correction in the absence of (fractional) cointegration.

As Clarke and Lebo (2003) show, to avoid bias and Type I errors, the ECM cannot be left as is—unlike Engle and Granger's model with a strictly stationary ECM, here the ECM may still have auto-correlation with $d > 0$. The next stage fractionally differences Y, X, *and* the ECM by each one's own d value, creating $\Delta^{d_Y} Y_t$, $\Delta^{d_X} X_t$, and the Fractional Error Correction Mechanism (FECM) ($\Delta^{d_{ECM}} ECM_t$) before the final step. For example, if Y and X are both $d = 0.8$ and the cointegrating residuals are $d = 0.5$, then Y and X need to be differenced by 0.8 and the ECM needs to be differenced by 0.5. The last step estimates

$$\Delta^{d_Y} Y_t = \alpha_0 + \alpha_1 \Delta^{d_{ECM}} ECM_{t-1} + \beta_1 \Delta^{d_X} X_t + \epsilon_t. \tag{8}$$

This follows the basic intuition of Engle and Granger's ECM framework and the logic of pre-whitening and balancing equations: both right- and left-hand-side variables must be free of

[33]This is not meant to be an exhaustive list of available techniques. For example, Brandt and Freeman (2006) evaluate several Bayesian and multi-equation approaches to time-series analyses in political science. For a thorough evaluation of Bayesian approaches to time series, see Bauwens, Lubrano, and Richard (2000).

[34]Two concerns are commonly raised over the estimation of d—the power of estimators with smaller samples and poor performance of estimators in the presence of significant short memory (Baillie and Kapetanios 2007). Byers, Davidson, and Peel (2000) investigate a wide range of public opinion series and find that the vast majority are well described as $(0,d,0)$ models; political time series rarely demonstrate the need for higher-order approximations (see also Lebo, Walker, and Clarke 2000). Estimates of the data sets we replicated affirm the conclusion of Byers et al., and we provide the Schwartz Bayesian criterion estimates for various ARFIMA (p, d, q) models in Appendix H.1. Regarding shorter time series, Robinson (1995) finds that the semi-parametric Whittle estimator is unbiased with as few as 64 observations. Grant (n.d.) investigates the finite sample properties of various estimators and finds that both semi-parametric and parametric estimators are unbiased with $(0,d,0)$ series, even with N as small as 40. Further, when higher-order approximations are necessary, the negative bias found in semiparametric estimators and the parametric exact MLE is not present in the frequency-domain approximate ML estimator. Appendix H.1 compares the long memory estimates of replicated data series using three estimators: the semi-parametric Whittle, the exact ML estimator, and the frequency-domain ML estimator. The estimates agree with each other closely.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 7 Results of FECM model with FI data, $T = 60$

Fractional order of integration (d=)	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*	*1.1*	*1.2*
FECM significant (%)	4.5	4.0	3.9	3.7	3.7	3.4	3.8	4.0
Significant X (%)	5.9	5.8	5.9	5.9	5.9	5.6	5.9	5.8
FECM significant with (%) $d < d(Y)$ and $d(X)$	1.7	1.9	1.9	1.9	2.1	1.9	2.3	2.4

Note. Percentage results in each cell based on 10,000 simulations per bivariate FECM model with null hypothesis of no fractional cointegration. Significance of FECM based on one-tail test. Significance of X based on two-tail test. DV and IV are fractionally integrated at the same level of d. $T = 60$.
FECM Model: $\Delta^{d_Y} Y_t = \alpha_0 + \alpha_1 \Delta^{d_{ECM}} ECM_{t-1} + \beta_1 \Delta^{d_X} X_t + \epsilon_t$.

auto-correlation to get trustworthy estimates (Hosking 1981; Tsay and Chung 2000; Lebo, Walker, and Clarke 2000; Clarke 2003). By transforming every component of the model to be (0,0,0), the three-step method establishes equation balance and protects inferences while being quite flexible.[35] Simulation results from FECM models with randomly generated FI series are shown in Table 7. The Type I error rate for error correction hypotheses stays below the 5% threshold regardless of the memory of the original series.

One disadvantage of the three-step model is interpretability. Explaining that "a one-unit change in fractionally differenced X leads to a β change in fractionally differenced Y" is difficult. And, while the FECM's α_1 coefficient in equation 8 has a similar interpretation to an ECM in Engle and Granger's two-step model—the speed of error correction—the coefficient will be slightly different from the value that relates to the level-form variables. In both cases, however, the extra difficulty in interpretability is a good trade-off for a trustworthy hypothesis test that avoids both Type I and Type II errors.

Pesaran et al.'s (2001) ARDL bounds test procedure falls between the GECM and fractional cointegration techniques in terms of flexibility:

$$\Delta Y_t = \beta_0 + \theta_1 Y_{t-1} + \theta_2 X_{t-1} + \sum_{i=1}^{p} \varpi_i \Delta Y_{t-i} + \sum_{j=0}^{p} \beta_j \Delta X_{t-j} + \epsilon_t. \tag{9}$$

The ARDL model uses the same ΔY_t as the GECM for the dependent variable, and the inclusion of the ΔY_{t-i} and ΔX_{t-j} terms on the right-hand side of the equation allows for serial correlation and ensure the error term is white noise. With the bounds testing approach, the regressors can be of mixed orders of integration—stationary, non-stationary, or fractionally integrated—and the use of bounds allow the researcher to make inferences even when the integration of the regressors is unknown or uncertain. The ARDL model allows greater flexibility on the right-hand side, but because it is a cointegration test, the dependent variable, Y_t, must still be $I(1)$. Further, the researcher must ensure that no series within the model is $I(2)$, the inclusion of which will invalidate the model's results.

The ARDL is a two-step procedure and the first step, bounds testing, refers to hypothesis testing the presence of a long-run relationship via an F-test, which tests the joint significance of the coefficients on the lagged level-form variables in equation 9, i.e., H_0: $\theta_1 = \theta_2 = 0$. The F-test has a non-standard distribution that depends on the number of regressors, the presence of an intercept or trend, and the sample size. The critical values for the bounds of the F-test were computed by PSS based upon sample sizes of 500 and 1000; however, the method is appropriate for use with smaller samples. For smaller samples between 30 and 80, Narayan (2005) provides proper critical values.

Should the F-test exceed the upper bound, the researcher can reject the null of no cointegration and conclude that a long-run relationship is present, irrespective of the order(s) of integration of the regressors. If the F-statistic falls below the lower bound, the null hypothesis cannot be rejected. Note that the presence of cointegration implies the presence of Granger causality in at least one

[35]The FECM model can be used in a near-VAR framework so that one can measure how multiple endogenous variables re-equilibrate to each other. See, for example, Lebo, McGlynn, and Koger (2007).

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 8 Dickey–Fuller results of CEW's dependent variables

Model	*All Reviews*	*Non-salient reviews*	*Salient reviews*
Dickey–Fuller coefficient	−0.18	−0.18	−0.51
Dickey–Fuller *t*-statistic	−2.28	−2.32	−3.98*
MacKinnon DF critical value	−2.94	−2.94	−2.94
Estimated *d* value	0.62 (0.11)	0.62 (0.12)	0.36 (0.08)

Note. Dickey–Fuller critical values are from MacKinnon (1994). *d* values estimated using Stata's exact ML estimator; Robinson's semi-parameteric estimator in RATS provides similar estimates: *All Reviews* ($d = 0.63$); *Non-Salient Reviews* ($d = 0.65$); *Salient Reviews* ($d = 0.30$).

direction; however, the ARDL does not indicate the direction of causal ordering (see Dickinson and Lebo [2007] for a short summary and example of the method). Among the model's disadvantages is the loss of degrees of freedom with longer lags. Additionally, the model's flexibility in terms of its allowance of uncertainty as to the regressors' orders of integration is helpful should the *F*-test exceed the bounds; however, if the computed *F*-statistic falls within the critical value band, the order(s) of integration must be determined before a correct conclusion can be made.

5 Applications

Next, we replicate five recent GECM studies. Two are presented below and three (Kelly and Enns 2010; Sanchez Urribarri et al. 2011; and Volscho and Kelly 2012) can be found in Appendices E.1, E.2, and E.3 of the Supplementary Materials. In each study, key findings change when using alternative methods.

5.1 *Example 1: Casillas, Enns, and Wohlfarth (AJPS, 2011)*

Casillas, Enns, and Wohlfarth (2011, CEW) investigate the effect of public mood on the Supreme Court's reversal of cases in a liberal direction and find a strong positive relationship for non-salient cases; for salient cases, public mood has no effect. These are compelling findings given the long debate on the relationship between public opinion and the Court's behavior. Using the GECM, CEW find Court responsiveness to be much stronger than previously shown. ECM parameters of −0.83, −0.77, and −1.27 are described as powerful equilibrium relationships where previous work had found tenuous ties at best.[36]

We first check the properties of the dependent variables.[37] As seen above, the further a series is from $I(1)$, the more negative the ECM coefficient and its *t*-statistic will be. As shown in Table 8, *All Reviews*, *Non-Salient Reviews*, and *Salient Reviews* test as fractionally integrated with the *d* value for *Salient Reviews* noticeably lower. Error correction is very likely to appear strong in a GECM of *Salient Reviews*, regardless of the covariates.

To demonstrate, we use CEW's DVs and simulate covariates to replicate their three models.[38] Table 9's Model 1 is based on 10,000 simulations per DV with three randomly generated (0,0,0) IVs. Following the authors' use of standard *t* critical values, the models have 66.3%, 66.5%, and 97.7% Type I error rates for error correction, respectively. Switching to MacKinnon values provides reasonable results for *All Reviews* and *Non-Salient Reviews*: no errors at all. That ΔX and X_{t-1} are significant far too often is still a problem considering the IVs were specifically generated as

[36]ECM parameters of All Review, Non-Salient Review, and Salient Review models, respectively. For Non-Salient Reviews, they report: "The error correction rate indicates that 77% of the long-run effect of mood will occur at term $t+1$ (0.68) and an additional 77% of the remaining effect will influence the Court at term $t+2$ (0.16). Therefore, 94% of the total long-run effect of public opinion at term t will be manifested in the justices' behavior after just two terms."

[37]Replications of CEW's models and further details are in Appendix C.

[38]We follow CEW's use of 2SLS but our findings are roughly the same using OLS (see Appendix C.3).

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 9 Simulations of results for CEW's dependent variables in two scenarios

	All reviews	*Non-salient reviews*	*Salient reviews*
Model 1[a]			
ECM Significant - one tail t-distribution (%)	66.3	66.5	97.7
ECM Significant - MacKinnon Values (%)	0	0	26.7
Mean of α_1	−0.16	−0.16	−0.51
Mean t-statistic of α_1	−1.79	−1.79	−3.39
ECM & $\geq 1\Delta X_t$ Significant (%)	19.1	16.4	24.3
ECM & $\geq 1X_{t-1}$ Significant (%)	15.4	12.1	23.7
Model 2[b]			
ECM Significant - one tail t-distribution (%)	94.7	95.2	99.9
ECM Significant - MacKinnon Values (%)	26.7	25.4	95.3
Mean of α_1	−0.45	−0.44	−0.96
Mean t-statistic of α_1	−3.23	−3.21	−5.87
ECM & $\geq 1\Delta X_t$ Significant (%)	26.6	23.5	31.3
ECM & $\geq 1X_{t-1}$ Significant (%)	53.9	51.8	78.5

Note. 10,000 simulations per model per DV. [a]Model 1: IVs are level stationary $I(0)$. [b]Model 2: IVs are unit-roots ($I(1)$). MacKinnon ECM critical values from Ericsson and MacKinnon (2002) for $T = 45$ with 3 IVs: −3.838.

approximations of white noise processes. The *Salient Reviews* column highlights the problems of estimating the GECM with a DV with a low order of fractional integration. The MacKinnon values find over 25% of ECMs significant and the average size of the α_1 parameter would be reported as strong re-equilibration.

Model 2 of Table 9 presents simulations for more realistic data—unit-root IVs. Even with MacKinnon values, we find a significant ECM over 25% of the time for *All Reviews* (26.7%) and *Non-Salient Reviews* (25.4%). The ECM coefficients, averaging −0.45 and −0.44, are far from zero and the hypothesis tests on ΔX_t and X_{t-1} are much worse—at least one of the three lagged level-form IVs is significant over 50% of the time. The *Salient Reviews* model is as bad as we would expect: the average ECM coefficient is −0.96, what CEW would call almost complete re-equilibration, with random data. Comparing the results of Model 1 and Model 2 shows that bias in α_1 as well as β_0 and β_1 increases markedly in relation to the memory of X_t, a reminder that pre-whitening is important for all variables in a model.

Model 2 also highlights inferential problems caused by the DV's order of integration. The Dickey–Fuller test fails to reject the null of a unit-root for both *All Reviews* and *Non-Salient Reviews*, but when regressing these DVs on simulated $I(1)$ series we reject the null of no error correction far too often, even with MacKinnon values. In this case, the fractional integration of the dependent variables confounds the hypothesis test.

Next, Table 10 shows the results when we replace CEW's three IVs with yearly *Worldwide Shark Attacks*, *U.S. Tornado Fatalities*, and *U.S. Beef Consumption*.[39] Using CEW's interpretation and favored t-value, we find strong evidence of error correction in all three models.[40] With an ECM of −1.04 (s.e. = 0.16), even using MacKinnon's critical values leads to finding extremely strong error correction in the *Salient Reviews* model.

[39]Beef Consumption (million tons) data from USDA, *Production, Supply, and Distribution* database (www.fas.usda.gov/psdonline); Shark Attack data from Florida Museum of Natural History (http://www.flmnh.ufl.edu/fish/sharks/statistics/Trends2.htm); Tornado Fatalities from NOAA (http://www.nws.noaa.gov/om/hazstats/resources/weather_fatalities.pdf).

[40]We have no evidence of endogeneity problems, but to match CEW we use a 2SLS model and instrument for *Beef Consumption* with CEW's excluded "social forces," which are used as instruments of Martin-Quinn scores. Social forces such as the homicide rate, policy liberalism, inequality, etc., may predict mood, but not MQ scores. Instrument tests for proper inclusion in the first stage fail with both our data and CEW's data. See Appendix C.2 for CEW's first-stage results and Appendix C.1 for the full replication of CEW.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 10 Using the GECM to explain the Court's liberal reversal rate with sharks, tornadoes, and beef consumption

	All reviews	*Non-salient reviews*	*Salient reviews*
Long-run multiplier			
Shark attacks	0.56*	0.53*	0.51*
	(0.20)	(0.18)	(0.19)
Tornado fatalities			
Beef consumption	−1.79*	−1.69*	−2.04*
	(0.32)	(0.78)	(0.31)
Long-run effects			
Shark attacks$_{t\text{-}1}$	0.30*	0.29*	0.53*
	(0.14)	(0.13)	(0.21)
Tornado fatalities$_{t\text{-}1}$	0.05	0.04	0.10
	(0.04)	(0.04)	(0.08)
Beef consumption (IV)$_{t\text{-}1}$	−0.95*	−0.92*	−2.12*
	(0.36)	(0.33)	(0.46)
Short run effects			
ΔShark attacks	0.20*	0.18	0.39*
	(0.12)	(0.11)	(0.22)
ΔTornado fatalities	0.06*	0.04*	0.10*
	(0.02)	(0.02)	(0.05)
ΔBeef consumption (IV)	−7.10*	−6.88*	−2.90
	(4.17)	(3.86)	(7.89)
Error correction and constant			
Percent liberal$_{t\text{-}1}$	−0.53*	−0.54*	−1.04*
	(0.17)	(0.17)	(0.16)
Constant	61.71*	61.52*	129.11*
	(21.75)	20.44	(25.58)
Fit and diagnostics			
Centered R^2	0.14	0.19	0.55
Sargan (χ^2)	0.69	0.60	0.32
N	45	45	45

Note. Entries are two-stage least squares coefficients (s.e. in parentheses). ECM significance one-tail *t*-test. Coefficient significance (*$p \leq 0.05$, one-tail test).

The ECM problems are compounded by spurious regressions—our nonsense IVs are significant far too often. Across all three models, seven of our nine short-term variables and six long-term variables—along with their respective LRMs—are significant.[41]

Because each of the three dependent variables used by CEW are estimated to be fractionally integrated, we also re-estimate their data using fractional methods. We first pre-whiten by fractionally differencing each series by its estimated order of fractional integration. The d estimates, standard errors, and confidence intervals for each series can be found in Appendix C, Table C.5. Each DV is fractionally integrated; however, we cannot reject the null that the three IVs are unit-roots. Because no DV / IV combination shares the same order of fractional integration, no model is a candidate for fractional cointegration.

Table 11 presents the results of regression models after all variables have been fractionally differenced, and what is immediately apparent is that the results are consistent with the short-term results reported by CEW. Pre-whitening caused no harm to the findings. Changes in either public mood or the ideological makeup of the Court will affect reversal rates. However, by ensuring

[41]The long-run multiplier (LRM) is calculated as the ratio of coefficients: β_1/α_1. The formula for the standard errors of the LRM is $((1/b^2)\mathrm{Var}(a) + (a^2/b^4)\mathrm{Var}(b) - 2(a/b^3)\mathrm{Cov}(a, b))^{1/2}$. See DeBoef and Keele (2008, 191–92).

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 11 Re-estimation with fractional methods (Casillas et al. 2011)

Review type	*Δ^d All reviews*	*Δ^d Non-salient reviews*	*Δ^d Salient reviews*
Short run effects			
Δ^d Public mood	1.43*	1.61*	0.90
	(0.73)	(0.72)	(1.76)
Δ^d Segal–Cover	9.61*	8.39*	14.51
	(4.67)	(4.58)	(11.30)
Δ^d Martin–Quinn	0.11	-0.30	1.79
	(2.30)	(2.26)	(5.56)
Constant	−1.35	−1.39	59.68*
	(1.25)	(1.23)	(3.02)
Fit and diagnostics			
Adjusted R^2	0.10	0.10	−0.02
Breusch–Pagan test (χ^2)	0.06	0.47	0.93
Breusch–Godfrey LM test (χ^2)	1.93	1.01	2.11

Note. Entries are OLS coefficients (s.e. in parentheses). All variables have been fractionally differenced by their estimate of *d*. Coefficient significance (*$p \leq 0.05$, two-tail test).

Table 12 ARDL bounds test for cointegration (Casillas et al. 2011)

	5% Critical bounds	
	I(0)	*I*(1)
	3.535	*4.733*
All review model	$F = 3.19$	
Non-salient review model	$F = 2.54$	

Note. *F*-statistic of unrestricted ARDL model with $K+1$ lagged-level form variables. Critical values from Narayan (2005) based on $T = 45$, $K = 3$, unrestricted intercept, and no trend. Neither model can reject the null of no cointegration.

equation balance, we avoided the spurious conclusions that an equilibrium relationship exists between the Court's reversal rate and public mood.

As a final step in our replication of this data, we also estimate the first step of the ARDL bounds test of Pesaran, Shin, and Smith (2001) with CEW's *All Review* and *Non-Salient Review* models. After estimating an unrestricted model, the *F*-statistic of the lagged-level variables was compared to the bounds computed by Narayan (2005). Recall from our simulations in Table 9, Model 2, that the GECM was susceptible to Type I errors due to the fact that both DVs are fractionally integrated. Despite this, we rely on the fact that neither DV can reject the null of a unit-root as justification for using the ARDL to test for a long-run relationship. The results are found in Table 12, and neither model's *F*-statistic exceeds the bounds. We cannot reject the null hypothesis of no cointegration, which indicates that the GECM is misspecified.

5.2 *Example 2: Ura and Ellis (Journal of Politics, 2012)*

Our second replication, (Ura and Ellis 2012, U&E), posits that macro-economic and political variables have asymmetric effects on partisan groups' responses and that this is at least partially responsible for mass polarization. U&E use General Social Survey items to create a measure of macro-policy sentiment among partisan groups. With $T = 35$, a large number of independent variables, and the use of the GECM, the paper's strong findings are suspect.

U&E's main results (their Table 2, replicated in Appendix D.1 of the Supplementary Materials) include error correction rates of−0.40 (s.e. = 0.09, $t = -4.54$) for Republicans and−0.69 (s.e. = 0.17, $t = -4.11$) for Democrats. Our Table 1 gives the appropriate critical value from Ericsson and

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

MacKinnon (2002) for $T=35$ as -4.268 for $I(1)$ series. Thus, there is some evidence of error correction and the adjusted-R^2 values of 0.39 indicate well-specified models.

But are the strong findings simply due to use of the GECM? We first replace their DVs—Republican and Democratic Mood—with randomly generated $I(1)$ series. The tendency to find significant error correction is strong—following U&E's use of a standard t-test, the ECM parameter is significant (0.05 level) in approximately 86% of the simulations.[42] Type I error rates on the other covariates are also troublesome. For example, in models where both the ECM and an X_{t-1} are significant, we observe equilibrating relationships in almost 62% of all simulations. Using MacKinnon values, we fare somewhat better but still make Type I errors on the ECM at over four times the rate we should (21.9% and 22.6%).

As discussed, MacKinnon values assume data are strictly $I(1)$ and become insufficient as any series in the model deviates from that. The correct 5% rejection region can only be found using the exact auto-correlation patterns of the data at hand.[43] In this case, 5% of the simulations have t-values below -5.81. Yet, even if it were feasible to ask researchers to run simulations to calculate idiosyncratic critical values, the bias in the ECM values and the Type I errors on the Xs make the GECM problematic with mixed orders of integration.[44]

Next, we try our Shark-Tornado-Beef data as predictors of U&E's dependent variables. Since U&E have five independent variables we add yearly values of *U.S. Onion Acreage* and *U.S. Coal Emissions*.[45] Table 13 shows the results: the ECMs in both equations are substantively large, and even with MacKinnon values we would claim that Republic Mood is in equilibrium with four of our five nonsense series. Many of the short-term and long-run independent variables are significant as well. With an R^2 of 0.57, the model of Republican Mood is noticeably better than U&E's and the Democratic model is just a touch worse ($R^2=0.29$). The ECM values, hypothesis tests, model fit statistics, and long-run multipliers all give us erroneous inferences.

Importantly, these are series with lots of memory—d varies between 0.89 and 1.44. Inattention to stationarity could potentially lead one to conclude that they are all simply $I(1)$ and that MacKinnon values will make for valid inferences. In fact, Dickey–Fuller tests on the *Domestic Spending* series indicates that it's $I(2)$, and inappropriate for use in either the GECM or ARDL models in its level form.

Finally, we fractionally difference U&E's variables and add a FECM between *Partisan Mood* and *Top 1% Income Share*, the predictor with the smallest p-value in U&E's models.[46] The FECMs in both models are indistinguishable from zero as are the independent variables in the *Democratic Partisan Mood* model.[47] Republicans, on the other hand, do appear to shift in a liberal direction following increases to *Defense Spending* and *Top 1% Income Share*. Thus, some of U&E's findings stand, but many do not hold up to these alternative specifications, and we find no evidence of error correction. As a final check, we run FECM models on our Shark-Tornado-Beef and find none of the FECMs and only one of 48 hypothesis tests on the independent variables to pass the 0.05-level of significance.[48] With $T=35$, estimation of d and fractional differencing are somewhat under-powered and some wariness is understandable. Still, the method is uncovering relationships we should expect (e.g., Defense Spending) and not setting off false alarms (e.g., Shark Attacks). With short series, we advocate the use of multiple-modeling techniques as checks for robustness (Table 14).

[42]Full simulation results are in Appendix D.1 of the Supplement.

[43]Estimates of d using Stata's exact ML estimator: DemMood ($d=1.15$, s.e. $=0.15$), RepMood ($d=1.05$, s.e. $=0.16$), Domestic Spending ($d=1.44$, s.e. $=0.08$), Defense Spending ($d=1.32$, s.e. $=0.11$), Inflation ($d=1.02$, s.e. $=0.25$), Unemployment ($d=0.94$, s.e. $=0.21$), Inequality ($d=0.89$, s.e. $=0.19$).

[44]Note that both *Defense* and *Domestic Spending* series are estimated as explosive processes, biasing the ECM t-statistic in the negative direction making Type I errors more likely. See Appendix A for further explanation. Additionally, the original DVs are bounded so our simulations of unbounded unit-root DVs are not exactly equivalent.

[45]Onion Data (10,000s of acres) from USDA (http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1396); Coal Emission (million tons) data comes from DOE (www.eia.gov/totalenergy/data/annual/showtext.cfm?t=ptb1101).

[46]We estimate FECMs between *Partisan Mood* and each of the other covariates and include those results in Appendix D.3 of the Supplementary Materials. All results are null.

[47]Running the model on first differences also returns null results.

[48]Table D.6 of Appendix D uses the independent variable with the lowest p-value (Onion Acreage) in the FECM. Each of the 4 other series takes a turn in the FECM in the models presented in Tables D.7 (Republican Mood) and D.8 (Democratic Mood).

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 13 What else moves mood? (Re-estimating Ura and Ellis (2012))

	Republican mood	*Democrat mood*	*|Difference|*
Long-run multipliers			
Onion acreage	−0.29* (0.04)	−0.20* (0.05)	0.09
Coal emissions	−0.13* (0.02)	−0.11* (0.02)	0.01
Beef consumption	−0.40 (0.26)	−0.03 (0.05)	0.36
Shark attacks	0.01* (0.00)	0.00 (0.00)	0.01
Tornado fatalities	−0.05* (0.01)	−0.02* (0.01)	0.03
Long-run effects			
Onion acreage$_{t-1}$	−0.18* (0.03)	−0.10* (0.02)	0.08*
Coal emissions$_{t-1}$	0.08* (0.01)	0.06* (0.01)	0.02
Beef consumption$_{t-1}$	0.25 (0.18)	0.02 (0.17)	0.23
Shark attacks$_{t-1}$	−0.01* (0.00)	0.00 (0.00)	0.01
Tornado fatalities$_{t-1}$	0.03* (0.01)	0.01 (0.01)	0.02*
Short-run effects			
ΔOnion acreage$_t$	0.01 (0.03)	0.03 (0.02)	0.02
ΔCoal emissions$_t$	0.08* (0.02)	0.05* (0.02)	0.04
ΔBeef consumption$_t$	0.83* (0.33)	0.20 (0.20)	0.63
ΔShark attacks$_t$	0.03* (0.02)	−0.00 (0.01)	0.03
ΔTornado fatalities$_t$	0.02* (0.00)	0.01 (0.01)	0.01*
Error correction and constant			
Partisan mood$_{t-1}$	−0.62* (0.11)	−0.52* (0.14)	0.11
Constant	12.08 (7.56)	20.52* (5.25)	
Fit and diagnostics			
Adjusted R^2	0.57	0.29	
Breusch–Pagan test (χ^2)	0.00	1.86	
Breusch–Godfrey LM Test (χ^2)	1.29	0.11	

Note. Entries are seemingly unrelated regression coefficients (s.e. in parentheses). ECM significance one-tail t-test. Coefficient significance (*$p \leq 0.05$, two-tail test). The difference column reports the absolute difference and t-test between coefficient estimates for the two models (*$p \leq 0.05$).

Table 14 Re-estimation with three-step FECM (Ura and Ellis (2012))

	Republican mood		*Democrat mood*	
Short run effects				
Δ^d Domestic spending \$10b	0.02	(0.04)	−0.00	(0.04)
Δ^d Defense spending \$10b	0.31*	(0.13)	0.11	(0.12)
Δ^d Inflation	−0.31	(0.20)	−0.05	(0.18)
Δ^d Unemployment	−0.47	(0.37)	−0.06	(0.32)
Δ^d Top 1% income share	1.22*	(0.42)	0.55	(0.37)
Error correction and constant				
FECM	0.12	(0.17)	-0.16	(0.19)
Constant	−1.30	(0.68)	−0.18	(0.59)
Fit and diagnostics				
R^2	0.38		0.11	
Breusch–Pagan Test (χ^2)	1.51		0.07	
Breusch–Godfrey LM Test (χ^2)	1.62		0.00	

Note. Entries are OLS coefficients (s.e. in parentheses). All variables have been fractionally differenced by their estimate of d. FECM significance (*$p \leq 0.05$, one-tail test). Coefficient significance (*$p \leq 0.05$, two-tail test).

6 Replication Roundup

Table 15 summarizes the results of our replications. The general pattern is an overstatement of significant relationships in the original papers. The body of work using the GECM would make it seem that error correction is a common pattern among political time series, but our reanalyses show

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Table 15 Findings and replications

Authors	*Central finding*	*Replication finding*
Casillas, Enns, and Wohlfarth (2011)	Public mood and Justice ideology exhibit significant short- and long-run effects on Court decision making. As public liberalism increases, so does the Court's rate of liberal decisions.	Public mood significant in short-term only. No evidence of cointegration for either *All Review* or *Non-Salient Review* models. *Salient Review* model inappropriate with mixed orders of integration.
Ura and Ellis (2012)	Partisan mood varies with political stimuli. Changes in mood and increases in polarization are asymmetric between parties with most movement coming from Republicans.	Short-term effects for defense spending and top 1% income share only. Three-step FECM better accounts for long-memoried series with no cointegration.
Sanchez Urribarri et al. (2011)	Institutional conditions and changes in judicial ideology explain the rights agendas of multiple high courts. No expected support for theory of exogenous structural change hypothesis.	Change in judicial ideology only significant for UK model. Insignificant in US model. No significant institutional effects. No support for exogenous structural change hypothesis. GECM model either inappropriate with mixed orders of integration or unnecessary given stationarity of the data.
Kelly & Enns (2010)	Income inequality exhibits both short-term and long-run effects on public mood. Both high-income and low-income populations exhibit increased conservatism in response to inequality.	No support for short- or long-term effect of income inequality on public mood. No support for either high or low income conservatism with increasing inequality. No cointegration - GECM model inappropriate.
Volscho & Kelly (2012)	Political, policy, and economic inputs all significantly affect income distribution of top 1%. Political inequality not only a result of market forces, but also a product of political change.	No support for politics or policy indicators having significant effect on *Top 1% Income*. Some economic indicators significant. Significant FECM between *Trade Openness* and *S&P Composite* with *Top 1% Income*. GECM model inappropriate with mixed orders of integration.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

otherwise. Error correction between variables is a very close relationship that should be obvious in a simple glance at the data. Non-intuitive findings of error correction should make political scientists highly suspicious.

One final lesson from these exercises is the value of using multiple approaches. Time-series analysts do not have the luxury of being able to replicate studies with new data, but robustness checks can come from finding similar results across diverse modeling choices.

7 Conclusion

Political scientists' troubles with the GECM are not surprising when the method is traced back to the econometrics literature, where its primary usage is as a cointegration test for $I(1)$ data (Banerjee et al. 1993; Ericsson and MacKinnon 2002). In that literature the GECM is not a one-step model, *it is the first step of a multi-step process*. The first step ignores the hypothesis tests on X_{t-1} and ΔX_t and uses MacKinnon values to test whether $\alpha_1 = 0$. Failing to reject that hypothesis means that cointegration is not present. Rejecting the null means that cointegration is present. In either case, it is not a standard regression model. If cointegration is not found, no ECM is specified. If it is found, additional steps follow. Re-postulating the GECM as a single-equation regression puts too many conditions on the model.

In fact, we recommend the GECM in only one rare situation: when all of the variables are strictly unit-root series, Y_t is unbounded, Y_t and X_t are cointegrated, and the MacKinnon critical values are used. We looked at many combinations of series' characteristics, and in every case but that hypothetical one, the GECM ran into serious problems.

A careful look at the applied literature in political science will not find any examples that meet all those criteria. When data are bounded or diverge from $I(1)$ by being stationary, auto-regressive, explosive, fractionally, or near-integrated, the ECM parameter begins to move dramatically. Adapting to MacKinnon critical values is not a sufficient fix, and generating idiosyncratic critical values is untenable as well. Bias in the ECM ruins its ability to do what it should—tell us about equilibrium relationships. As for the GECM's use with stationary data, the ADL and GECM may be mathematically equivalent but the GECM adds complications without adding useful insights. If Y's long-term tendency is simply to revert to its mean, we should not prefer a model whose purpose it to find Y's long-run relationship with X.

As every parent knows, the solution to the babies and bathwater dilemma is not to just leave the baby sitting in the dirty water—the two can be carefully separated. With time series, the safe solution is to return to the logic of pre-whitening and use a multi-step process that removes auto-correlation from each variable prior to including it in a multivariate model. Time-series data are complicated—too complicated to provide reliable results in a simple single step.

References

Baillie, Richard T., and Tim Bollerslev. 1994. Cointegration, fractional cointegration, and exchange rate dynamics. *Journal of Finance* 49(2):737–45.

Baillie, Richard T. 1996. Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73(1):5–59.

Baillie, R., and George Kapetanios. 2007. Semiparametric estimation of long memory: The holy grail or a poisoned chalice. In *International Symposium on Financial Econometrics*, pp. 1–2.

Banerjee, Anindya, Juan Dolado, and Ricardo Mestre. 1998. "Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework." *Journal of Time Series Analysis* 19(3):267–283.

Banerjee, Anindya, Juan J. Dolado, John W. Galbraith, and David Hendry. 1993a. Co-integration, error correction, and the econometric analysis of non-stationary data. *OUP Catalogue*.

———. 1993b. *Integration, error correction, and the econometric analysis of non-stationary data*. Oxford: Oxford University Press.

Bårdsen, Gunnar. 1989. Estimation of long run coefficients in error correction models. *Oxford Bulletin of Economics and Statistics* 51(3):345–350.

Bauwens, Luc, Michel Lubrano, and Jean-Francois Richard. 2000. *Bayesian inference in dynamic econometric models*. New York: Oxford University Press.

Beck, Nathaniel. 1992. Comparing dynamic specifications: The case of presidential approval. *Political Analysis* 3:27–50.

Beran, Jan. 1994. *Statistics for long-memory processes*, vol. 61. Boca Raton, FL: CRC Press.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Bernhard, William, and David Leblang. 2002. Democratic processes, political risk, and foreign exchange markets. *American Journal of Political Science* 46:316–33.

Box, George E.P., and David A. Pierce. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65(332):1509–26.

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

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

———. 1998. Investigating political dynamics using fractional integration methods. *American Journal of Political Science* 42:661–89.

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

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.

Büthe, Tim, and Helen V. Milner. 2014. Foreign direct investment and institutional diversity in trade agreements: Credibility, commitment, and economic flows in the developing world, 1971–2007. *World Politics* 66(1):88–122.

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. 2005. Limited time series with a unit root. *Econometric Theory* 21:907–45.

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 Kon S. Lai. 1993. A fractional cointegration analysis of purchasing power parity. *Journal of Business & Economic Statistics* 11(1):103–12.

Clarke, Harold D., and Marianne C. Stewart. 1994. Prospections, retrospections and rationality: The "Bankers" model of presidential approval reconsidered. *American Journal of Political Science* 38:1104–23.

Clarke, Kevin A. 2003. Nonparametric model discrimination in international relations. *Journal of Conflict Resolution* 47(1):72–93.

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.

Davidson, James E.H., David F. Hendry, Frank Srba, and Steven Yeo. 1978. Econometric modelling of the aggregate time-series relationship between consumer's expenditure and income in the United Kingdom. *Economic Journal* 88:661–92.

DeBoef, Suzanna. 2001. Modeling equilibrium relationships: Error correction models with strongly autoregressive data. *Political Analysis* 9(1):78–94.

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

———. 1999. Testing for cointegrating relationships with near-integrated data. *Political Analysis* 8(1):99–117.

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

Dickinson, Matthew J., and Matt 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. Of forests and trees. *Political Analysis* 4:255–58.

———. 1993. What moves policy sentiment? *American Political Science Review* 87:158–70.

Durr, Robert H., Andrew D. Martin, and Christina Wolbrecht. 2000. Ideological divergence and public support for the Supreme Court. *American Journal of Political Science* 44:768–76.

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

———. 2010. *Applied econometric time series*. 3rd 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.

Enns, Peter K. 2014. The public's increasing punitiveness and its influence on mass incarceration in the United States. *American Journal of Political Science* 58:857–72.

Enns, Peter K., Nathan J. Kelly, Janna Morgan, Thomas Volscho, and Christopher Witko. 2014. Conditional status quo bias and top income shares: How U.S. political institutions have benefitted the rich. *Journal of Politics* 76:289–303.

Ericsson, Neil R., and James G. MacKinnon. 2002. Distributions of error correction tests for cointegration. *Econometrics Journal* 5(2):285–318.

Faricy, Christopher. 2011. The politics of social policy in America: The causes and effects of indirect versus direct social spending. *Journal of Politics* 73(1):74–83.

Fuller, Wayne A. 2009. *Introduction to statistical time series*, vol. 428, Wiley.com.

Granger, Clive W.J. 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14(2):227–38.

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

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

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.

Granger, C.W.J. 2010. Some thoughts on the development of cointegration. *Journal of Econometrics* 158:3–6.

Grant, Taylor, and Matthew J. Lebo. 2016. Replication data for: Error correction methods with political time series. Harvard Dataverse, V1 http://dx.doi.org/10.7910/DVN/VAKZ2G.

Hansen, Bruce E. 1995. Rethinking the univariate approach to unit root testing. *Econometric Theory* 11(5):1148–71.

Hendry, David F., and G.J. Anderson. 1977. *Testing dynamic specification in small simultaneous systems: An application to a model of building society behaviour in the United Kingdom*. New Haven, CT: Yale University.

Hendry, David F., Frank Srba, and Stephen Yeo. 1978. Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal* 88(352):661–92.

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

Jennings, Will, and Peter John. 2009. The dynamics of political attention: Public opinion and the queen's speech in the United Kingdom. *American Journal of Political Science* 53(4):838–54.

Kayser, Mark Andreas. 2009. Partisan waves: International business cycles and electoral choice. *American Journal of Political Science* 53(4):950–70.

Keele, Luke. 2007. Social capital and the dynamics of trust in government. *American Journal of Political Science* 51(2):241–54.

Kelly, Nathan J., and Peter K. Enns. 2010. Inequality and the ynamics of public opinion: The self-reinforcing link between economic inequality and mass preferences. *American Journal of Political Science* 54(4):855–70.

Kono, Daniel Yuichi. 2008. Democracy and trade discrimination. *Journal of Politics* 70(4):942–55.

Kremers, Jeroen J.M., Neil R. Ericsson, and Juan J. Dolado. 1992. The power of cointegration tests. *Oxford Bulletin of Economics and Statistics* 54(3):325–48.

Layman, Geoffrey C., Thomas M. Carsey, John C. Green, Richard Herrera, and Rosalyn Cooperman. 2010. Activists and conflict extension in American party politics. *American Political Science Review* 104(2):324–46.

Leblang, David, and William Bernhard. 2006. Parliamentary politics and foreign exchange markets: The world according to GARCH. *International Studies Quarterly* 50(1):69–92.

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.

Lebo, Matthew J., and Will H. Moore. 2003. Dynamic foreign policy behavior. *Journal of Conflict Resolution* 47(1):13–32.

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 Everett Young. 2009. The comparative dynamics of party support in Great Britain: Conservatives, Labour and the Liberal Democrats. *Journal of Elections, Public Opinion and Parties* 19(1):73–103.

MacKinnon, James G. 1994. Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business & Economic Statistics* 12(2):167–76.

Narayan, Paresh Kumar. 2005. The saving and investment nexus for China: Evidence from cointegration tests. *Applied Economics* 37(17):1979–90.

Nicolau, Joao. 2002. Stationary processes that look like random walks: The bounded random walk process in discrete and continuous time. *Econometric Theory* 18(1):99–118.

Ostrom, Charles W. Jr., and Renée Smith. 1993. Error correction, attitude persistence, and executive rewards and punishments: A behavioral theory of presidential approval. *Political Analysis* 4:127–84.

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.

Phillips, A William. 1954. Stabilization policy in a closed economy. *Economic Journal* 64(254):290–323.

———. 1957. Stabilization policy and the time-forms of lagged responses. *Economic Journal* 67(266):265–77.

Pickup, Mark. 2009. Testing for fractional integration in public opinion in the presence of structural breaks: A comment on Lebo and Young. *Journal of Elections, Public Opinion and Parties* 19(1):105–16.

Ramirez, Mark D. 2009. The dynamics of partisan conflict on congressional approval. *American Journal of Political Science* 53(3):681–94.

Rickard, Stephanie J. 2012. Welfare versus subsidies: Governmental spending decisions in an era of globalization. *Journal of Politics* 74(4): 1171–83.

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

Sanchez Urribarri, Raul A., Susanne Schorpp, Kirk A. Randazzo, and Donald R. Songer. 2011. Explaining changes to rights litigation: Testing a multivariate model in a comparative framework. *Journal of Politics* 73(2):391–405.

Sargan, John Denis. 1964. Wages and prices in the United Kingdom: A study in econometric methodology. *Econometric Analysis for National Economic Planning* 16:25–54.

Stimson, James A. 1991. *Public opinion in America: Moods, cycles, and swings*. Boulder, CO: Westview Press.

Stock, James H., and Mark W. Watson. 2011. Introduction to econometrics, 3/E. *Instructor* 10.

Tsay, Wen-Jay, and Ching-Fan Chung. 2000. The spurious regression of fractional integrated processes. *Journal of Econometrics* 96:155–82.

Ura, Joseph Daniel. 2014. Backlash and legitimation: Macro political responses to Supreme Court decisions. *American Journal of Political Science* 58(1):110–26.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press

Ura, Joseph Daniel, and Christopher R. Ellis. 2012. Partisan moods: Polarization and the dynamics of mass party preferences. *Journal of Politics* 74(1):277–91.

Ura, Joseph Daniel, and Patrick C. Wohlfarth. 2010. "An appeal to the people": Public opinion and congressional support for the Supreme Court. *Journal of Politics* 72(4):939–56.

Volscho, Thomas W., and Nathan J. Kelly. 2012. The rise of the super-rich power resources, taxes, financial markets, and the dynamics of the top 1 percent, 1949 to 2008. *American Sociological Review* 77(5):679–99.

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.

Yule, G. Udny. 1926. Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society* 89(1):1–63.

Zivot, Eric. 2000. The power of single equation tests for cointegration when the cointegrating vector is prespecified. *Econometric Theory* 16(3):407–39.

https://doi.org/10.1093/pan/mpv027 Published online by Cambridge University Press