The General Error Correction Model in Practice

Matthew J. Lebo, PhD
Patrick W. Kraft, PhD Candidate

Stony Brook University

Forthcoming at
Research and Politics

May 3, 2017

Abstract

Enns, Kelly, Masaki, and Wohlfarth (2016) respond to recent work by Grant and Lebo (2016) and Lebo and Grant (2016) that raises a number of concerns with political scientists' use of the general error correction model (GECM). While agreeing with the particular rules one should apply when using unit root data in the GECM, Enns et al. still advocate procedures that will lead researchers astray. Most especially, they fail to recognize the difficulty in interpreting the GECM's "error correction coefficient." Without being certain of the univariate properties of one's data it is extremely difficult (or perhaps impossible) to know whether or not cointegration exists and error correction is occurring. We demonstrate the crucial differences for the GECM between having evidence of a unit root (from Dickey-Fuller tests) versus actually having a unit root. Looking at simulations and two applied examples we show how overblown findings of error correction await the uncareful researcher.

Introduction

In a recent symposium in *Political Analysis* Grant and Lebo (2016) and Lebo and Grant (2016) raise a number of concerns with use of the general error correction model (GECM). In response, Enns et al. (2016, EKMW) have contributed "Don't jettison the general error correction model just yet: A practical guide to avoiding spurious regression with the GECM." EKMW are prolific users of the GECM; separately or in combination they have authored 18 publications that rely on the model, often relying on significant error correction coefficients to claim close relationships between political variables. In "Don't jettison..." the authors narrow the gap of disagreement between themselves and G&L. However, as they attempt to reconcile old findings with new insights, EKMW inadvertently make clear an essential point: using the GECM is more complicated in practice than researchers realize. Despite their extensive experience, EKMW are still misinterpreting the inferences provided by the error correction coefficient and as a result are overstating relationships between variables.

In this paper we explain where we agree with and diverge from EKMW. In short, there is agreement that: a) with stationary data (I(0)) the GECM's parameters have different meaning and the strong possibility of user error makes the model a poor choice, and b) the GECM is more easily interpretable with all unit root (I(1)) and jointly cointegrated data so long as one uses the correct critical values. Many disagreements remain. In particular, despite the weaknesses of the Dickey and Fuller (1979) stationarity test, EKMW treat the test's results as perfectly reliable for identifying unit roots. We show both the high frequency of the Dickey-Fuller test misclassifying series as unit roots and the consequences for using such series in the GECM. Further, EKMW advocate stretching the use of "unit root rules" into other data scenarios but ignore the possible consequences of doing so.

We also explore differences in our understanding of the data used in Kelly and Enns (2010) and Casillas, Enns, and Wohlfarth (2011). EKMW argue that the data and analyses in those papers and potentially many others fit neatly into the unit root rules category. We maintain that EKMW are likely misclassifying the series in those papers as unit roots which can lead to over-stated claims of error correction. We begin with the points of agreement between EKMW and Grant and Lebo (2016).

1 Points of Agreement

1.1 The GECM can work when all the series contain unit roots and are jointly cointegrated

The GECM has several representations but the one most commonly used by political scientists is DeBoef and Keele's (2008, D&K) Equation 5:

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

EKMW and G&L agree with the literature in econometrics that when both X and Y contain unit roots – defined as $y_t = 1 * y_{t-1} + \epsilon_t$ – and are cointegrated, the combination of Y_{t-1} and X_{t-1} is stationary and the equation is balanced. In such cases Equation 1 has acceptable Type I error rates for all its parameters and is readily interpretable.

In particular, there is agreement that when both X and Y contain unit roots:

1. α_1^* functions as a test of cointegration between X and Y and measures the rate of error correction, theoretically bounded between 0 and -1.

2. The critical values for $\hat{\alpha}_1^*$ are non-standard, more negative than with the normal distribution, vary with the number of Xs, vary with the length of the data, and can be calculated as "MacKinnon values" following Ericsson and MacKinnon (2002).

3. When $\hat{\alpha}_1^*$'s t-statistic does not surpass the MacKinnon critical value there is no cointegration. Since Y_{t-1} and X_{t-1} are not in combination stationary there is unresolved autocorrelation on the right-hand-side and the model's estimates should not be used.

This is progress. G&L point out in their Table 2 how frequently a researcher might mistakenly conclude error correction is present if she were to incorrectly use the normal distribution to evaluate α_1^* with unit root series. EKMW recognize this when they say (p. 3): "Thus, the bottom row of Grant and Lebo's Table 2 should be read as evidence of the importance of using the correct MacKinnon critical values when testing for cointegration, not evidence of spurious relationships with the GECM." To be sure, G&L's Table 2 is one of many of their analyses intended to demonstrate what happens if – as Kelly and Enns (2010) and Casillas, Enns, and Wohlfarth (2011) do – one uses common but incorrect practices.[1]

1.2 The GECM is possible *but not recommended* when all series are stationary

EKMW and G&L agree on another key point: the GECM must be interpreted differently when all the data are stationary compared to when they all contain unit roots.

DeBoef and Keele (2008) and Keele, Linn, and Webb (2016) explain the equivalence of the GECM (Equation 1 above) to the ADL (Equation 2):

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

Stationary series require the "stationary rules" for Equation 1: 1) α_1^* does not test cointegration, 2) Y_{t-1}'s hypothesis test evaluates $\hat{\alpha}_1^* + 1$ and MacKinnon values are not used (Bannerjee, Dolado, Galbraith, and Hendry 1993, p. 167), and 3) estimates must be translated to the ADL framework as $\alpha_1^* + 1 = \alpha_1$, $\beta_0^* = \beta_0$, and $\beta_1^* = \beta_0 + \beta_1$. Thus, when $\hat{\alpha}_1^* = -1.00$ with a $(0, 0, 0)$ series it indicates stationarity – no impact of Y_{t-1} on Y_t in the ADL.

We did not find these post-estimation calculations in any of our selected readings of the roughly 500 papers that cite DeBoef and Keele (2008). Instead, when data are claimed to be stationary, the value and significance of $\hat{\alpha}_1^*$ and $\hat{\beta}_1^*$ are taken from software output and treated the same as they would be using the unit root rules. Typically, this leads to

[1] To our knowledge, political scientists had not used MacKinnon values with the GECM before G&L.

overconfidence in rejecting null hypotheses and in finding error correction to be occurring.

Thus, G&L do not say that the GECM *cannot* be used with stationary data, but argue (p.4): "...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." In other work Kelly, Enns, and Wohlfarth did not adapt their interpretation of the GECM while arguing data are stationary but, with "Don't Jettison the GECM Just Yet," the authors are now on board, saying (p. 6): "Thus, we agree with Grant and Lebo that when the dependent variable is stationary, the parameterization of the GECM is more likely than the ADL to lead to errors of *interpretation*."

This is also progress. DeBoef and Keele (2008) advocate the GECM with stationary data – an early version of the paper was entitled "Not Just for Cointegration: Error Correction Models with Stationary Data."[2] With all stationary series, they argue, one can estimate an ECM without discussing cointegration, long-run equilibria, or error correction rates. However, in addition to interpretation problems, other issues followed as well.

A key misreading of D&K is to conclude that the GECM is perfectly flexible so that series of any type can be analyzed together within it.[3] In particular, D&K's statement (p. 199) that "...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" has been repeatedly quoted but seldom understood.[4] The applied literature is peppered with statements such as: "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" (Volscho and Kelly 2012), "The ECM provides a conservative empirical test of our argument and a general model that is appropriate with both stationary and nonstationary data" (Casillas, Enns, and Wohlfarth

[2]D&K argued that since the ADL and GECM are simply reparameterizations of each other and since the ADL works with stationary data, then the GECM must as well.

[3]The misreading of D&K has also led to reduced emphasis on the univariate properties of data. This has always been a crucial step in time series analysis but much of the GECM literature ignores the question of stationarity of data or gives it short shrift.

[4]See G&L Appendix F for misinterpretations of this line in the applied time series literature.

2011), and "While the use of an ECM is often motivated by the presence of a nonstationary time-series as a dependent variable, our application of this model is based on the fact that it is among the most general time-series models that imposes the fewest restrictions" (Kelly and Enns 2010). Engle and Granger's (1987) strict rules for cointegration were increasingly ignored as the GECM became the dominant technique in political science.

EKMW's conclusion (p. 10) that "Although the ADL and GECM produce the same information (in different formats), the ADL is less likely to yield errors of *interpretation* when Y is stationary" matches G&L's (p.27): "...with stationary data, the ADL and GECM may be mathematically equivalent but the GECM adds complications without adding useful insights." For example, Casillas, Enns, and Wohlfarth (2011) use the obviously stationary series *Salient Reviews* in the GECM and report an error correction rate of 126%, precisely the kind of misinterpretation the ADL can avoid. Thus, while all agree on the mathematical facts, from a practical standpoint EKMW and G&L are on one side of the issue – recommending against using the GECM with stationary data – and D&K are on the other.

2 Point of Likely Disagreement

On another point agreement is uncertain. G&L posit that the univariate properties of *all* series in the GECM deserve attention; e.g. if all the independent variables are I(1) they must all be cointegrated with the dependent variable. As opposed to Engle & Granger's (1987) two-step ECM or Clarke and Lebo's (2003) three-step fractional ECM, the GECM does not allow testing for cointegration or measuring error correction between Y and a subset of Xs.[5]

For example, the cointegration of unit root series Y and X in Equation 1 makes the component $(Y_{t-1} + X_{t-1})$ jointly stationary and, along with ΔY_t and ΔX_t, all components will then be stationary and inferences can be carefully drawn. Adding an I(1) X_2 means adding two predictors, $\Delta X_{2,t}$ and $X_{2,t-1}$ but if X_2 is not jointly cointegrated with Y and

[5]For example, Lebo and Young (2009) test for cointegration between vote intentions and leadership approval ratings for each of Britain's three major parties. Two- and three-step approaches model error correction between the two without specifying it between vote intentions and economic indices. Lebo, McGlynn, and Koger (2007) estimate error correction between democratic unity, republican unity, and democratic size in Congress but leave other independent variables for the full regression model.

X the model creates problems due to unresolved autocorrelation in $X_{2,t-1}$. Thus, even if cointegration exists between Y and X, researchers need to be more concerned about the properties of other Xs. EKMW (p.9) do not seem worried, for example, defending Casillas *et al*'s Table 1 and Table 2 (model 1) even though both include a Social Forces variable that is not significant in either lags or differences.[6] More generally, the consequences of additional non-stationary Xs that are not cointegrated are not well understood but are often included in GECM applications. Next we turn to areas of more explicit disagreement.

3 Points of Disagreement

To review, all agree that a bivariate GECM estimates parameters α_0, α_1^*, β_0^*, and β_1^* and that with unit root data we evaluate each "as is" but use MacKinnon values for the ECM parameter, α_1^*. Also, with stationary data we need to switch the rules: $\beta_1^* = \beta_0 + \beta_1$ of the ADL, α_1^* is not a cointegration test, and $\alpha_1^* + 1$ relies on the t-distribution.

Our views deviate as we confront the stark choice about which rules to apply, especially with respect to the error correction coefficient. When should we switch from one set of rules to the other? EKMW claim that it is possible to unambiguously choose rules based on results from augmented Dickey-Fuller (ADF) tests (p. 4): "If the ADF rejects the null of a unit root, we do not use the GECM to test for cointegration."

However, Dickey-Fuller tests have a null hypothesis of a unit root so that positive evidence is required to classify the series as not I(1). As EKMW admit (p.4) "it is well known that ADF tests are underpowered against the alternative hypothesis of stationarity" meaning that many series incorrectly show evidence of a unit root. Indeed, the ADF test is sensitive to sample size, trends, and bounds. Fractionally integrated, near-integrated, autoregressive, and other stationary series often fail to reject the null in the ADF test. ADF tests can also be affected by trending, periodicity, and heteroskedasticity.

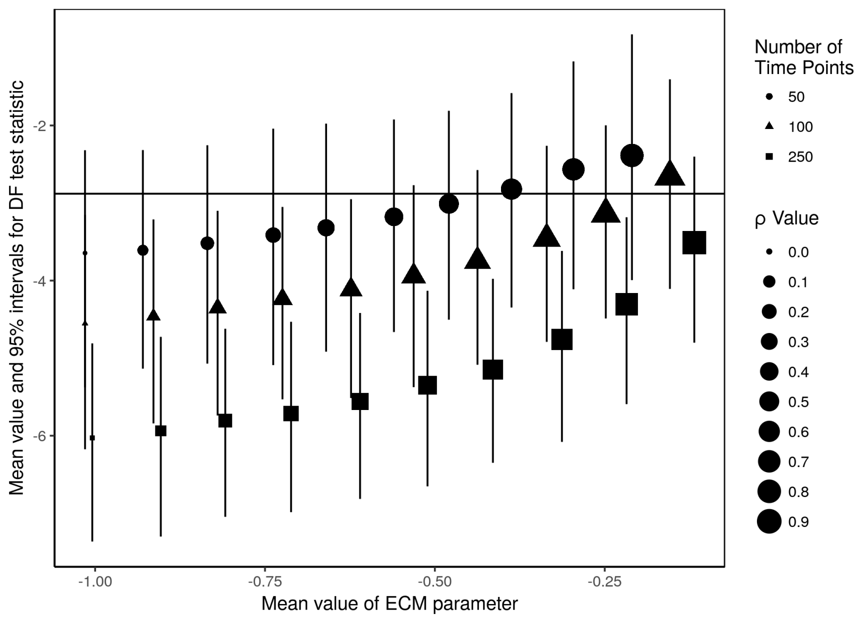
Falsely classifying series as I(1) means being too quick to favor the GECM over the ADL, to apply the wrong rules to the GECM, and to think that lower values of α_1^* indicate error

[6] Social Forces is part of an instrumental variable analysis where the instrument tests are not passed.

correction between series and not simply the stationary tendencies of Y.

Figures 1(a), 1(b), and 1(c) show these problems for the GECM. We generated 60,000 pairs of unrelated time series – 10,000 each for T=50, T=100, and T=250 while varying ρ and then again while varying d. Figure 1(a) shows, for each T, 1,000 pairs of simulated autoregressive (ρ) series for each of $(0,0,0)$ to $(0.9,0,0)$ in increments of 0.1. Figure 1(b) shows 1,000 pairs of fractionally integrated series simulated as $(0,0,0)$ up to $(0,0.9,0)$ increasing d in increments of 0.1. Thus, none of these series contain a unit root.[7]

Figure 1(a): With $\rho < 1$, ADF false negatives are rampant and occur with downward biased α_1^*



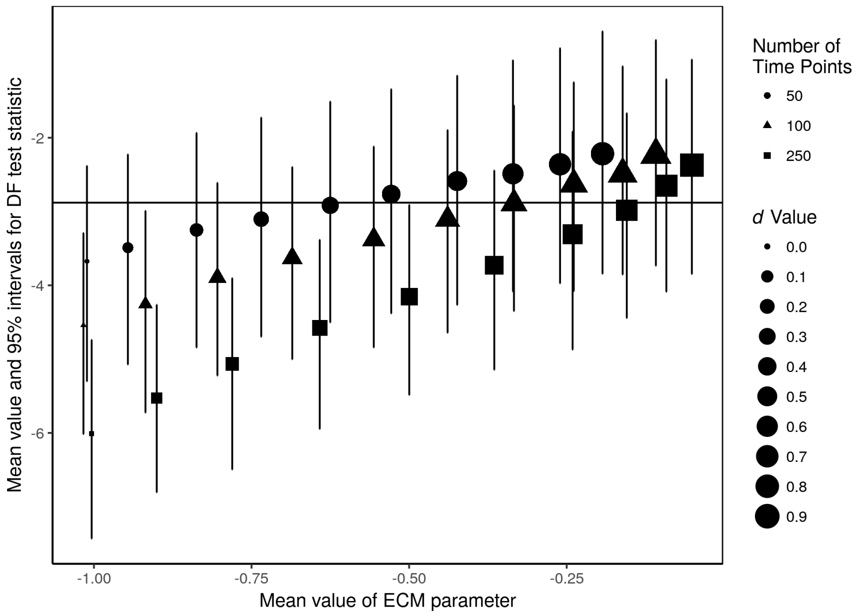
Note: Each data point represents 1,000 simulations of a bivariate GECM with a particular autoregressive parameters. All points above the horizontal line represent Type II errors (false negatives) with the Dickey-Fuller unit root test.

For both Figures 1(a) and 1(b), each shape shows the average value of 1,000 ADF test statistics of Y with vertical whiskers showing coverage of 950 of the 1,000 statistics. The horizontal line is the .05-level critical value above which are false negatives – a failure to reject the I(1) null with data that are not I(1). The many instances above the line indicate that the test is drastically underpowered.[8]

[7] We do not include analyses for more complicated ARFIMA processes with p and q parameters but such series are possible and would add further confusion to understanding ADF test results.

[8] With true unit roots the DF test also has problems. When we simulate I(1) data with sample sizes of

Figure 1(b): With $d < 1$, ADF false negatives are rampant and occur with downward biased α_1^*



Note: Each data point represents 1,000 simulations of a bivariate GECM with a particular fractional integration parameter. All points above the horizontal line represent Type II errors (false negatives) with the Dickey-Fuller unit root test.

On the X-axis is the average estimated $\hat{\alpha}_1^*$ from the GECM. As ρ or d moves away from I(1) the ECM value drops lower, seemingly – but not actually – indicating error correction.

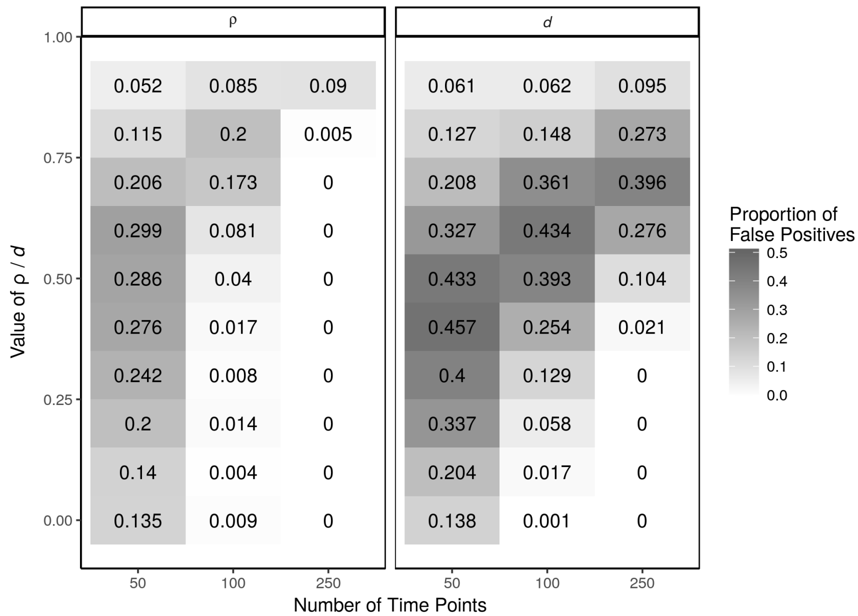
The figure's results should be striking, most especially for short time series. Unrelated series simulated as $(0, 0.5, 0)$ with T=50 have an average ECM value of -0.53 while failing to reject the ADF null 61.1% of the time. Series that are $(0.5, 0, 0)$ with T=50 have an average ECM value of -0.56 but appear to be unit roots in the DF test 36.9% of the time.[9] When $d = 0.8$ and T=50 – about where many yearly public opinion series fall – the ADF test has false negatives at a rate of 76.1%. In these cases we'd find an average ECM of -0.26 and be well on our way to touting error correction. With longer series there are still problems.

Of course, the high rate of false negatives on the ADF test would not be as problematic if the ECM parameter testing for cointegration (α_1^*) did not reach statistical significance.

50, 100, and 250 we incorrectly reject the null hypothesis 14.3%, 17.1%, and 19.2% of the time, respectively.

[9]Since there are many versions of the ADF test an over-zealous researcher might try several in search of one that provides evidence of a unit root. Thus, in practice, the probability of a false negative across the range of DF tests is actually much higher.

Figure 1(c): Proportion of false positives on α_1^* using MacKinnon critical values *coinciding with* ADF false negatives; i.e. falsely finding cointegration.



Note: Based on the same simulations as Figures 1(a) and 1(b). Each cell represents 1,000 simulations of a bivariate GECM. The data are constructed to hold a range of properties for variation in the autoregressive and fractional integration parameters. The percentages indicate the proportion of simulations where the Dickey Fuller unit root test failed to reject and the ECM parameter (α_1^*) is significant based on MacKinnon critical values.

Proponents of the GECM like EKMW might suggest that using MacKinnon critical values for the ECM parameter would prevent us from finding false evidence of cointegration in such a scenario. Unfortunately, MacKinnon critical values are not a panacea here, since they rely on the assumption of unit root data and are not extreme enough to prevent falsely finding error correction when series are not unit roots.

Figure 1(c) shows this. Each cell reports the proportion of times that the ADF test fails to reject a false null hypothesis of a unit root *and* the ECM parameter is significant beyond MacKinnon critical values. For example, with series created as $(0.6, 0, 0)$ and T=50 there is a 29.9% chance of concluding both that Y has a unit root and that it is cointegrated with X. With series created as $(0, 0.6, 0)$ and T=100 the rate is 43.4% for finding cointegration when following the exact procedures EKMW advocate. The problems are noticeably more pronounced with data we create as fractionally integrated compared to autoregressive. Additionally, shorter time series are much more problematic – at T=250 the problems remain for fractionally integrated series with higher values of d but disappear when there is only autoregression and the ADF test is more powerful.

To reiterate, not a single series in Figures 1(a), 1(b), or 1(c) contains a unit root. Thus, the GECM's unit root rules should not be used for any of them.[10] Still, EKMW's favored ADF test will incorrectly conclude that many are I(1). In fact, with short series the ADF test gives us false negatives a majority of the time and even 15% of the time *when series are complete white noise*, i.e. $(0, 0, 0)$. Nevertheless, EKMW would advise applying the unit root rules and MacKinnon critical values to those ADF false negatives without realizing that much of the apparent error correction – see $\hat{\alpha}_1^*$ decreasing from right to left – is due to the distance the series is from actually being I(1).[11] In all, the figures show how easy it is to have faulty evidence of both a unit root and error correction.

EKMW do not confront these problems. Instead, they force non-unit root data into the "all unit root" case and rely on the error correction parameter (α_1^*) for key inferences. EKMW interpret the GECM's results in the same way whether their series have "evidence of a unit root" or are actually simulated as unit roots.[12] In reality, data are messy and many non-unit root series will provide evidence of being I(1). So, yes, understanding α_1^* with data simulated to be exact unit roots is straightforward but this does not mean we can reliably interpret the coefficient when using real world time series with unknowable properties.

If we could identify with certainty I(1) series we could know when to apply the unit root rules but, as EKMW correctly point out (p. 9), "It may be that with short time series, we cannot draw firm conclusions about the time series properties of variables." Given that, EKMW should not choose rules based on a weak test and should not focus on a parameter

[10] Although the consequences of making this mistake when ρ or d are nearly 1 might be minimal (De Boef and Granato 1999).

[11] For example, G&L tested Casillas, Enns, and Wohlfarth (2011)'s *All Reviews* as $(0, 0.62, 0)$. Following Figure 1(a), such series fail to reject the null hypothesis 69.3% of the time and give an average ECM value of -0.42 even with an unrelated independent variable.

[12] The data in G&L's and EKMW's Case 1 simulations are unit roots. Reexamining Casillas et al.'s data EKMW report (p. 9): "Yet, the balance of evidence from the various tests suggest that these series contain a unit root."

whose interpretability gets muddier as data deviate from exactly I(1). Doing so gambles with inferences since if the data are not truly I(1) then α_1^* does not mean what they think it means.

Elsewhere, EKMW are explicit about extending the unit root rules to data that are not I(1). In their Case 4, EKMW apply them to near-integrated data – where ρ is close to but not equal to one in $y_t = \rho * y_{t-1} + \epsilon_t$.[13] Such series may provide evidence of a unit root but are technically mean stationary. How do we choose which set of rules to use? Banerjee, Dolado, Galbraith, and Hendry (1993) (p. 225) say in the context of ECMs: "In finite samples the differences between, for example, an AR(1) with parameter 1.0 and an AR(1) with parameter 0.99 is a difference of degree rather than kind." So perhaps we should not switch rules when $\rho = 0.99$ just because the series is technically stationary.

But the unit root rules are not exactly correct when $\rho = 0.99$ either. As Figures 1(a), 1(b), and G&L's Table 4 show, as data move away from unit roots there is a steady progression from 0 to -1 in the estimation of $\hat{\alpha_1^*}$. *This means that the correct critical values are even more extreme than MacKinnon's values.* We could derive correct idiosyncratic critical values if we could simulate data with the exact same properties but this is a practical impossibility.

Thus, when do we switch from one set of rules to the other? There is no magic threshold as a series goes from $\rho = 1.00$ to $\rho = .99$ or from $\rho = .90$ to $\rho = .89$ where on one side α_1^* is a cointegration test and the error correction rate and on the other side it is neither. At some point an ADF test statistic will tip from non-significant to significant but this cannot tell us the extent to which α_1^* speaks to error correction. Using more extreme MacKinnon values prevents many false positives in EKMW's simulation exercises but it does not mean they are correct when data are not unit roots.[14]

[13] EKMW discuss near-integrated series as being common in political science but this is not supported by the applied literature.

[14] G&L say (p. 15): "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

EKMW oversimplify again when they use unit root rules for fractionally integrated (FI) series where $0 < d < 1$.[15] Unsurprisingly, they find that many spurious results can be avoided by applying MacKinnon values to G&L's FI simulations and say (p.7): "Again we find, however, that the different conclusions can be resolved by following Grant and Lebo's advice to test for cointegration with the correct critical values."

This seems disingenuous. Neither G&L, Ericsson and MacKinnon (2002), nor any other source we know of has argued that MacKinnon values are appropriate except with exact I(1) data. EKMW provide no justification for expanding when these values should be used – to NI data, FI data, or any other type. Yes, EKMW's advice prevents some spurious findings but that does not mean that these are the correct critical values. As G&L's Figure 4 shows, α_1^*'s distribution quickly gets even more extreme than MacKinnon's distribution as d decreases from 1. Unless we simulate data ourselves, we cannot be sure of the exact (p, d, q) models which means we cannot calculate exactly what the idiosyncratic critical values are. G&L point out that "Even if we could pin down the correct critical values, the meaning of the ECM coefficient has been lost. Ultimately, the value of α_1 tells us more about the level of memory in Y_t than about Y_t's relationship to independent variables in the model."

In sum, many series that do not have unit roots will test as though they do. Also, neither fractionally integrated nor near-integrated series have unit roots and thus do not work for the critical values set out in Ericsson and MacKinnon (2002).[16] Treating such series as I(1) in the GECM means α_1^* will move downwards as the series deviate from I(1) – too often surpassing the MacKinnon values EKMW would like to use more liberally. Researchers that

distribution – they are unique to the particular data being used."

[15] Esarey (2016) sums up his contribution to the *Political Analysis* symposium with: "Based on these conclusions, it appears that ADL/ECM models are very useful for recovering the immediate impact of x on y, despite fractional integration. The results for long-run impacts are not quite as robust: these impacts are likely to be incorrectly estimated by an ADL/ECM run on fractionally integrated data." Esarey's simulations use only data where $0 < d < 0.5$. When $1.0 < d < 0.5$ (where studies find many political time series (Lebo, Walker, and Clarke 2000)) the problems of interpreting α_1^* will worsen. Helgason (2016) finds fractional integration methods to perform as well as or better than the GECM when series are characterized as $(0, d, 0)$ and that long-run effects are better investigated using FI methods even with samples as small as T=50.

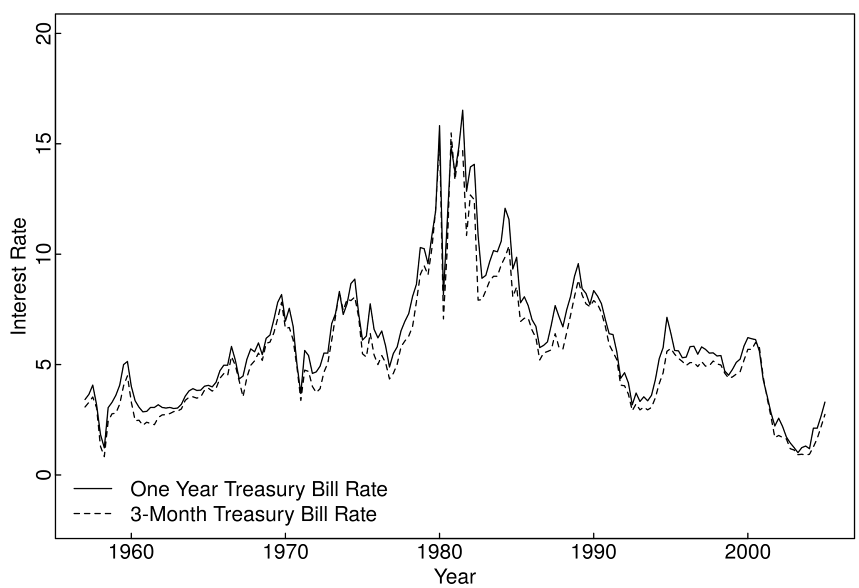
[16] As DeBoef and Granato (1997, p. 619) say: "both (near and fractionally integrated data) make characterizing the nature of our data more difficult by blurring the knife-edge distinctions between integrated and stationary processes."

mistakenly treat non-I(1) series as I(1) will misstate the meaning of α_1^*. Thus, EKMW's advice to apply MacKinnon values to estimates of α_1^* when Y is fractionally integrated, near-integrated, or fails to reject the ADF null invites incorrect claims of cointegration and error correction.

4 The GECM in practice when we are too quick to find unit roots

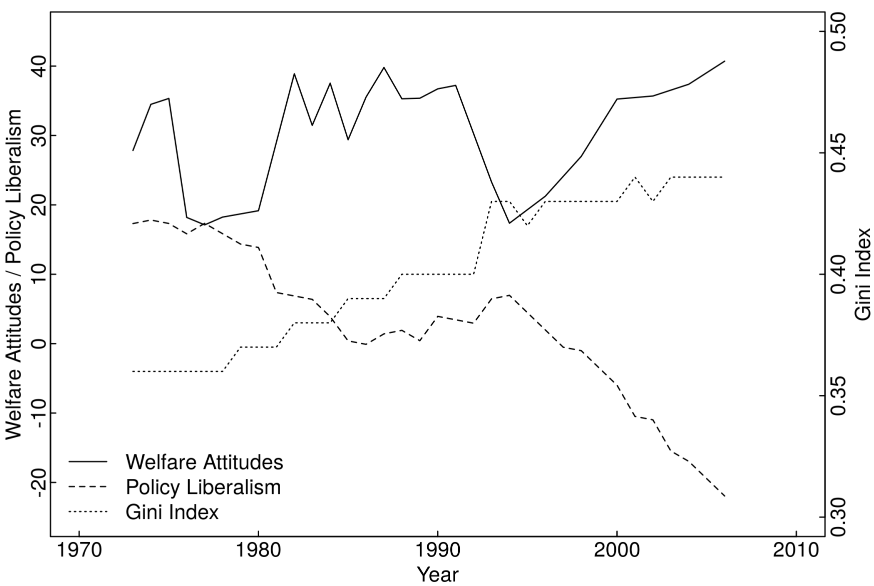
Next we consider the practical implications of squeezing non-unit root data into the unit root case for the GECM. To begin, recall Murray (1994)'s story that a unit root variable is like a drunk out for a walk – the next step is random but his current location is the sum of the steps taken thus far. Cointegration is akin to the drunk taking a leashed dog along for the walk. The two may be on random walks but are tethered so that any distance between them is eventually closed (error correction) and in the long run tends to zero.

Figure 2: Stock and Watson's cointegration example: Three-month and one-year T-bill rates, error correction rate=52%



What data do economists study for error correction? The textbook example in Stock and Watson (2011) uses one-year and three-month treasury bill rates set by the federal reserve, shown in Figure 2. These are unit roots; unless the Fed decides to change them – a shock in the error term – interest rates at time t are what they were at $t-1$. The relationship between the rates appears very close and the GECM indeed shows cointegration with $\alpha_1^* = -0.52$. That is, 52% of a gap between the series at t is closed at $t+1$ and 52% of the remaining gap is closed at $t+2$ and so on. How do political scientists' stories compare?

Figure 3: Kelly and Enns' data in Table 1 Model 4; K&E, EKMW claim error correction rate is 55%



4.1 Another look at Kelly and Enns (2010)

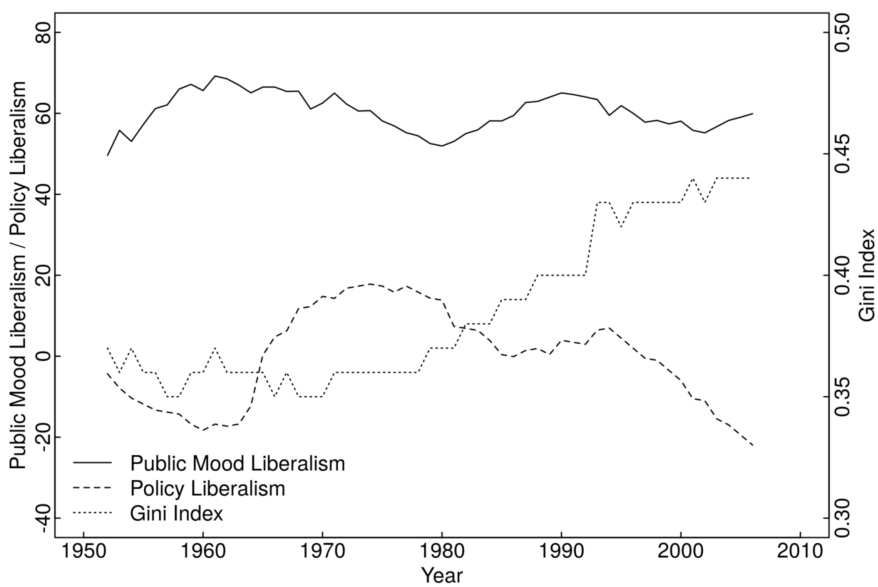
EKMW defend Kelly and Enns's (2010, K&E) results so long as the unit root rules are applied. K&E's Table 1 Model 4 shows the GECM results with Welfare Attitudes as the dependent variable and Policy Liberlism and the GINI index as independent variables. With just T=33 it is unsurprising that ADF tests on all three variables fail to reject the null of a unit root.[17] Thus, the data surpass EKMW's threshold to apply the unit root rules.[18]

[17] The 5% critical value is -2.978. Welfare test statistics are -1.879 (0 lags), -2.046(1 lag), -2.005(2 lags); the Gini Index test statistics are 0.017 (0 lags), 0.537(1 lag), 0.821 (2 lags), -2.443 (2 lags and trend with CV=-3.498); Policy Liberalism test statistics are 0.010 (0 lags), -0.929 (1 lag), -1.337 (2 lags).

[18] The construction of public opinion time series such as Welfare Attitudes make them unlikely to be I(1). The value at t may be highly correlated with $t-1$ but values are generated anew at each time point and do not exhibit random walk behavior. That is, we should not believe that absent some shock Y_t is exactly equal to Y_{t-1}. Studies have found series like these to be fractionally integrated (see, e.g., Box-Steffensmeier and Smith 1996; Byers, Davidson, and Peel 2000; Gil-Alana 2003; Box-Steffensmeier, De Boef, and Lin 2004;

Figure 3 shows these series. The solid line plots Y, Welfare Attitudes, and the two dotted lines are the Xs.[19] With T=33, rejecting the unit root hypothesis of the ADF test is very unlikely and the mean-reverting tendencies of Y are affecting the estimation of $\hat{\alpha_1^*}$. By classifying the series as unit roots, EKMW call a significant $\hat{\alpha_1^*}$ evidence of cointegration – i.e. the series in Figure 3 are tethered together. In fact, EKMW insist K&E's reported error correction rate of 55% is correct. That is, *K&E's claims in their AJPS article rest on our believing that the error correction relationship in Figure 3 is stronger than in Figure 2.*[20] Rather, the figures should be convincing that EKMW are misinterpreting their results.

Figure 4: Data from Kelly and Enns's Table 1 Model 2 - relationships are not apparent



Note: The GECM's estimated error correction rate is -0.25.

Elsewhere, EKMW (p. 4) specifically defend models in K&E and say "Yet, looking at Kelly and Enns' most parsimonious analysis (Table 1, column 2) we find clear evidence of

Lebo and Cassino 2007; Lebo 2008).

[19]The GECM does not allow us to specify which X we think Y is error correcting with so it could be that K&E think Welfare is cointegrated with the GINI coefficient, Policy Liberalism, or both. If all the variables are I(1) they must all be cointegrated or the model is misspecified.

[20]The 5% MacKinnon value for T=35 and 2 Xs is -3.613 and the test statistic on K&E's ECM is -3.46. EKMW might report this as evidence of cointegration at the .1 level or they might say the lack of cointegration means the model should be discarded. In any event, the MacKinnon values are not appropriate.

cointegration." The three series are graphed in Figure 4 with Public Mood Liberalism (the solid line) as the dependent variable. It is possible that cointegration is hard to see when more than two series are involved but making the comparison between K&E's data and the classic example in Figure 2 it seems more likely that the error correction rate is overstated by K&E – Figure 4 does not look like a drunk and her dog(s).

Also, compare K&E's Models 1 and 2 in their Table 1 which show the same error correction rate ($\hat{\alpha}_1^* = -0.25$, s.e.=0.07). The t-statistics of the six covariates in Model 1 are 1.48, -1.85, -0.01, 0.14, -0.64, and -0.46. Why is $\hat{\alpha}_1^*$ exactly the same in the two models - one of which has no Xs that matter? Because the models have the same Y and when they interpret $\hat{\alpha}_1^*$ EKMW are confusing Y's mean reverting behavior with error correction between Y and X.[21] Falsely inferring error correction is an easy mistake to make and shows the risks of relying on inferences drawn from $\hat{\alpha}_1^*$. Researchers are on safer ground when they concentrate on inferences drawn from the βs and long-run multipliers.

4.2 Another look at Casillas, Enns, and Wohlfarth (2011)

Next, we look at EKMW's defense of Casillas, Enns, and Wohlfarth (2011). CEW use three dependent variables: Salient Reviews, Non-Salient Reviews, and All Reviews decided in a liberal direction at the U.S. Supreme Court. EKMW (note 24) say:

> "We agree with Grant and Lebo that Casillas, Enns, and Wohlfarth were wrong to interpret the t-statistic on the lagged value of *salient* reversals as evidence of cointegration. This series is stationary, [...] so cointegration and long-run relationships should not have been considered."

What EKMW miss, however, is that the differences between Salient Reviews on the one hand and All Reviews and Non-Salient Reviews on the other are ones of degree, not category. These variables are computed anew each year based on the Court's decisions, making them

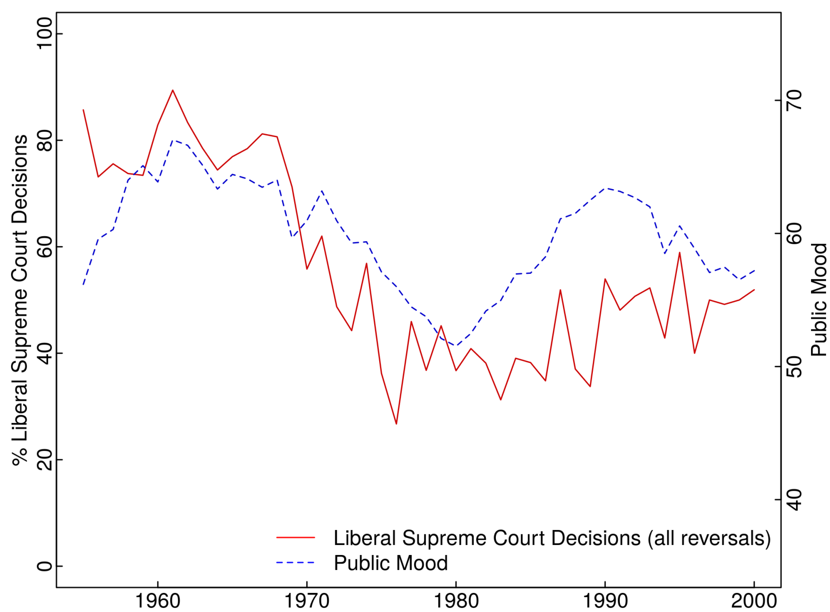
[21] K&E's four models in their Table 2 report error correction rates of -0.45, -0.46, -0.58, and -0.57 but these might vary based simply on the persistence of the dependent variables.

very unlikely to contain unit roots. However, with T=45, ADF tests have extremely low power in confirming that.

EKMW stand by CEW's estimates for All Reviews ($\hat{\alpha_1^*} = -0.83$) and Non-Salient Reviews ($\hat{\alpha_1^*} = -0.77$). Describing their Table 1, CEW say (p. 80):

> "The significant long-run impact of mood on the Court suggests that public opinion also has an effect that is distributed over future time periods. The error correction rate of 0.83 indicates the speed at which this long-term effect takes place. We expect that 83% of the long-run impact of public mood will influence the Court at term $t + 1$ (0.72), an additional 83% of the remaining effect will transpire at term $t + 2$ (0.12), and so on until the total long-run effect has been distributed. Therefore, the Courts long-term responsiveness to public mood occurs rather quickly, as 97% of the total long-run effect of public opinion at term t will be manifested in the justices behavior after just two terms."

Figure 5: Casillas et al.'s data for Table 1, error correction rate claimed to be 83%



This seemingly incontrovertible conclusion (t=-5.33) flies in the face of the well established attitudinal model (Segal and Spaeth 2002) but is based on short data and a parameter that is difficult to understand. Figure 5 plots out All Reviews and Public Mood. The series look more related than K&E's data and there may indeed be a close relationship there. However, CEW's claim that the error correction rate is 83% implies a much faster rate than what is presented in the T-bill example above.[22] Comparing the figures should make it clear that CEW and EKMW are exaggerating error correction – their α_1^* estimates may be capturing mean reversion or, perhaps, *both* mean reversion *and* the long run effects of X.

It is impossible to know the extent to which a negative coefficient on Y_{t-1} simply indicates that if the level of the series was high (or low) in the last period ΔY will be negative (or positive) in the present period due to mean reversion. A significant $\hat{\alpha_1^*}$ can imply different things but it is extremely difficult to distinguish among them. Without complete confidence that the dependent variable has a unit root, judging the extent of error correction in $\hat{\alpha}_1^*$ is unknowable with current GECM techniques. Even MacKinnon values are not extreme enough to prevent Type I errors.

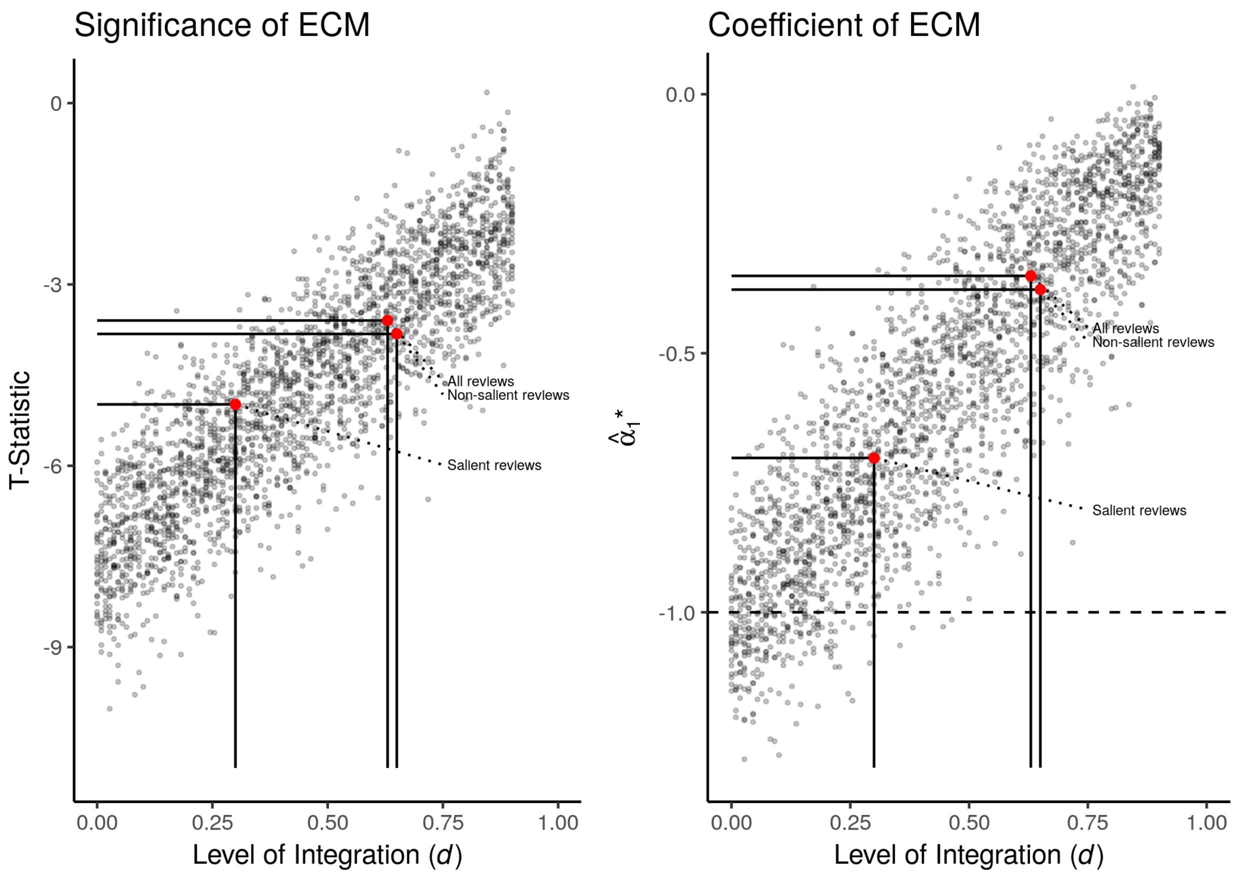
To demonstrate, Figure 6 begins with data from G&L's simulations of fractionally integrated data and shows decreasing values of d associated with increased values and t-statistics for $\hat{\alpha_1^*}$. Of the many dots, only those on the extreme right of each panel contain I(1) dependent variables but, as shown in Figure 1(b), many others would provide I(1) evidence.

Figure 6 also includes $\hat{\alpha_1^*}$ estimates from bivariate GECM models for CEW's DVs and Public Mood. EKMW classify *Salient Reviews* as stationary and admit the unit root rules should not be applied. But, with ADF results that *Non-Salient Reviews* and *All Reviews* are unit roots, they apply the unit root rules and find long-run relationships. However, the low $\hat{\alpha}_1^*$ value and t-statistics are due – at least to some extent – to Y's stationary tendencies. G&L estimate $d = 0.62$ for both – over 3 standard errors below one.[23] Overlaying the CEW results by the d estimates shows the findings fall exactly where they would be if no

[22]For the data in Figure 2 to correct 97% of the distance between the series it would take 5 periods, compared to the 2 periods described by CEW for gaps between All Reviews and Public Mood.

[23]Even if G&L's d estimates are off a bit (Lebo and Weber 2015; Grant 2015) it is implausible to think the Supreme Court's percentage of liberal decisions this year begins with last year's percentage and adds this year's shocks. The series does not have a unit root.

Figure 6: Casillas et al.'s ECM estimates are where they would be if no error correction was occuring



Note: The ECM coefficients for Casillas et al.'s three dependent variables are plotted alongside the simulated data from Grant and Lebo's (2006) Figure 3. The strength of error correction is a clear function of the level of integration in both the simulated data and all three of CEW's dependent variables.

relationship exists between X and Y.

In all, EKMW's statement (p. 2): "...we reconsider two of the articles that Grant and Lebo critiqued (CEW, K&E) and we demonstrate that a correct understanding of the GECM indicates that the methods and findings of these two articles are sound." could only be true if the *weakest* relationship among our Figures 2, 3, 4, and 5 is Figure 2's textbook example of cointegration. EKMW's blanket statement also misdirects from the fact that they only defend those papers' least outrageous findings.

In fact, CEW misinterpret $\hat{\alpha}_1^*$ and have no real evidence that "the public mood directly constrains the justices' behavior and the Court's policy outcomes, even after controlling for the social forces that influence the public and the Supreme Court." The long run equilibrium the Supreme Court series is reverting to may just be its own mean, not the public's mood. Finally, even if the independent variables are determining $\hat{\alpha}_1^*$, CEW's approach does not allow them to isolate which x is constraining the Court – the Court's ideology is an equally likely explanation.

In practice, if you are using the unit root rules in the GECM, you can only easily interpret $\hat{\alpha_1^*}$ if you are certain the dependent variable has a unit root. This is a near impossible task unless one is simulating the data, the time series is quite long, or one has a deep understanding of the data generating process. Given near certain uncertainty, it is best to either find a different model or to rely on other inferences the model provides.

5 Further Thoughts on Simulations and Replications

EKMW (p.2) say: "Although our conclusions differ greatly from Grant and Lebo's recommendations, we do not expect our findings to be controversial. Most of our evidence comes directly from Grant and Lebo's own simulations." This statement deserves more attention than there is space for here but the essential point is that EKMW's widespread promotion of MacKinnon values is an easy way to reduce Type I error rates in simulations or in practice *but these are not the right critical values except when data truly have a unit root.* When d or ρ is very close to one MacKinnon values may be close but they are not correct. When d or ρ are further from 1 tests often mistakenly find unit roots.

The properties of the series EKMW use – in K&E, CEW, and elsewhere – do not match the properties of the data they simulate. The consequences of being wrong are to get nonsense results, e.g. insisting that 97% of the disequilibrium between Supreme Court decisions and public mood is corrected in two years.

Also, EKMW cast doubt when they say G&L misused the GECM by replacing the independent variables of published work with shark attacks, tornado fatalities, other nonsense series, and simulated data. EKMW are correct that in G&L's replications they did not follow their own advice to set aside regression results where there is no evidence of cointegration. But, of course, *that was precisely the point*: G&L are demonstrating the mistakes made when GECM results are misinterpreted as was done by K&E, CEW, and the many published GECM studies.

That is, G&L show that if one mimics the methods and interpretation of papers like Casillas, Enns, and Wohlfarth (2011) the independent variables do not really matter in terms of getting significant results on the error correction parameter.[24] EKMW are correct, for instance, that G&L's Table 10 replication of CEW would not show shark attacks and beef consumption to be related to Supreme Court decisions if G&L had improved upon CEW's methods. Nevertheless, it is notable that with both Ura and Ellis's (2012) and Kelly and Enns' (2010) data, some relationships – G&L Table 13 (Republican Mood) and G&L's Table E.13 Model 3, respectively – between the DVs and variables such as Onion Acreage are strong enough to surpass MacKinnon values and conclude cointegration exists. Since the data are unlikely to have unit roots, MacKinnon values still are not enough to prevent spurious findings of error correction.

6 Conclusion

We show that using the unit root rules with series that simply pass the Dickey-Fuller test is not enough to avoid overstating findings of error correction. Especially with short time series it is too easy to fail to reject the DF null of non-stationarity and then misunderstand the error correction coefficient. This is a principal reason why the applied literature is replete with incorrect GECM findings and why G&L recommend against using the model except in ideal circumstances.

EKMW's "Don't Jettison the GECM" tries to clarify how to correctly interpret the ECM coefficient but inadvertently shows that while the GECM is viewed as flexible and easy to

[24]In their Appendix 3 EKMW complain that G&L simulate only the independent variables in their Tables E.11, E.12, and E.13 and argue that: "If one (or more) series do not vary across simulations, valid inferences cannot be made." G&L's point is that the univariate properties of a DV can lead to problematic results. If regressing the DV on simulated data and nonsense series like shark attacks leads to many Type I errors, the point is supported.

use it is, in fact, inflexible and extremely easy to misuse. EKMW agree that the GECM can work when data are unit roots, cointegrated, and special critical values are used but fail to realize they cannot simply extend those rules to non-unit root data and obtain reliable conclusions.

Indeed, EKMW advocate applying the unit root rules to data they know to not have unit roots – autoregressive and fractionally integrated – as well as other series that merely show evidence of a unit root. Moreover, they make their decisions based on the much maligned Dickey-Fuller test. Although they say (p. 2): "Most of our evidence comes directly from Grant and Lebo's own simulations" EKMW do so while roughly doubling the critical values without proving the practice is correct. Many spurious findings are eliminated this way but MacKinnon values and the unit root rules are not enough to overcome the interpretation problems on α_1^* when data are not I(1).

G&L say (p. 27): "Error correction between variables is a very close relationship that should be obvious in a simple glance at the data." The graphs of data we provide here should make it clear that EKMW's claims to have found long run equilibria between political time series in K&E and CEW come from the misuse of the method, not the data. Finding long run equilibria across decades of data makes for a story that is interesting but that hinges on a parameter that is often inscrutable. If researchers insist on employing the GECM with data that may not be unit roots they need to focus on the model's other parameters.

References

Banerjee, Anindya, Juan J Dolado, John W Galbraith, and David Hendry. 1993. "Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data." *OUP Catalogue* .

Bannerjee, Anindya, Juan Dolado, John W. Galbraith, and David F. Hendry. 1993. *Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data.* Oxford: Oxford University Press.

Box-Steffensmeier, Janet M., and Renée M. Smith. 1996. "The Dynamics ofAggregate Partisanship." *American Political Science Review* 90: 567–580.

Box-Steffensmeier, Janet M., Suzanna De Boef, and Tse-Min Lin. 2004. "The Dynamics of the Partisan Gender Gap." *American Political Science Review* 98: 515–528.

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.

Clarke, Harold D., and Matthew Lebo. 2003. "Fractional (co)integration and Governing Party Support in Britain." *British Journal of Political Science* 33(02): 283–301.

De Boef, Suzanna, and Jim Granato. 1999. "Testing for Cointegrating Relationships with Near-Integrated Data." *Political Analysis* 8(1): 99–117.

DeBoef, Suzanna, and Jim Granato. 1997. "Near-integrated Data and the Analysis of Political Relationship." *American Journal of Political Science* 41: 619–640.

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

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimatorsfor Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74: 427–431.

Engle, Robert F., and C. W. J. Granger. 1987. "Co-integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55: 251–276.

Enns, Peter K, Nathan J Kelly, Takaaki Masaki, and Patrick C Wohlfarth. 2016. "Don't jettison the general error correction model just yet: A practical guide to avoiding spurious regression with the GECM." *Research & Politics* 3(2): 2053168016643345.

Ericsson, Neil R, and James G MacKinnon. 2002. "Distributions of Error Correction Tests for Cointegration." *The Econometrics Journal* 5(2): 285–318.

Esarey, Justin. 2016. "Fractionally integrated data and the autodistributed lag model: Results from a simulation study." *Political Analysis* 24(1): 42–49.

Gil-Alana, Luis A. 2003. "Testing of Fractional Cointegration in Macroeconomic Time Series*." *Oxford Bulletin of Economics and Statistics* 65(4): 517–529.

Grant, Taylor. 2015. "Fractional integration in short samples: Parametric versus semiparametric methods." *Available at SSRN 2615098* .

Grant, Taylor, and Matthew J. Lebo. 2016. "Error correction methods with political time series." *Political Analysis* 24(1): 3–30.

Helgason, Agnar Freyr. 2016. "Fractional integration methods and short time series: Evidence from a simulation study." *Political Analysis* 24(1): 59–68.

Keele, Luke, Suzanna Linn, and Clayton McLaughlin Webb. 2016. "Treating time with all due seriousness." *Political Analysis* 24(1): 31–41.

Kelly, Nathan J, and Peter K Enns. 2010. "Inequality and the Dynamics of Public Opinion: The Self-Reinforcing Link Between Economic Inequality and Mass Preferences." *American Journal of Political Science* 54(4): 855–870.

Lebo, Matthew J. 2008. "Divided government, united approval: the dynamics of Congressional and Presidential approval." *Congress and the Presidency* 35(2): 1–16.

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): 464–481.

Lebo, Matthew J., and Christopher Weber. 2015. "An Effective Approach to the Repeated Cross-Sectional Design." *American Journal of Political Science* 59: 242–258.

Lebo, Matthew J., and Daniel Cassino. 2007. "The Aggregated Consequences of Motivated Reasoning and the Dynamics of Partisan Presidential Approval." *Political Psychology* 28(6): 719–746.

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.

Lebo, Matthew J, and Taylor Grant. 2016. "Equation balance and dynamic political modeling." *Political Analysis* 24(1): 69–82.

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.

Murray, Michael P. 1994. "A drunk and her dog: an illustration of cointegration and error correction." *The American Statistician* 48(1): 37–39.

Segal, J.A., and H.J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited.* Cambridge University Press.

Stock, James H., and Mark W. Watson. 2011. *Introduction to Econometrics*. 3rd ed. Addison-Wesley.

Ura, Joseph Daniel, and Christopher R Ellis. 2012. "Partisan Moods: Polarization and the Dynamics of Mass Party Preferences." *The Journal of Politics* 74(1): 277–291.

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): 679–699.