ORIGINAL ARTICLE

Detecting true relationships in time series data with different orders of integration

Peter K. Enns[1], Carolina Moehlecke[2] and Christopher Wlezien[3]* 🆔

[1]Cornell University, Ithaca, NY, USA, [2]Fundação Getulio Vargas, São Paulo, SP, Brazil and [3]University of Texas at Austin, Austin, TX, USA
*Corresponding author. Email: wlezien@austin.utexas.edu

Abstract

It is fairly well-known that proper time series analysis requires that estimated equations be balanced. Numerous scholars mistake this to mean that one cannot mix orders of integration. Previous studies have clarified the distinction between equation balance and having different orders of integration, and shown that mixing orders of integration does not increase the risk of type I error when using the general error correction/autoregressive distributed lag (GECM/ADL) models, that is, so long as equations are balanced (and other modeling assumptions are met). This paper builds on that research to assess the consequences for type II error when employing those models. Specifically, we consider cases where a true relationship exists, the left- and right-hand sides of the equation mix orders of integration, and the equation still is balanced. Using the asymptotic case, we find that the different orders of integration do not preclude identification of the true relationship using the GECM/ADL. We then highlight that estimation is trickier in practice, over finite time, as data sometimes do not reveal the underlying process. But, simulations show that even in these cases, researchers will typically draw accurate inferences as long as they select their models based on the observed characteristics of the data and test to be sure that standard model assumptions are met. We conclude by considering the implications for researchers analyzing or conducting simulations with time series data.

Keywords: Dynamic modeling/analysis; time series models

Equation balance is of paramount importance in time series analysis. An unbalanced equation is misspecified and prone to type I error; that is, the rejection of a true null hypothesis. Although this is well known to statisticians, there is confusion in political science about what constitutes balance. Indeed, in his contribution to a recent symposium on time series analysis in *Political Analysis*, John Freeman (2016, 50) wrote, "It now is clear that equation balance is not understood by political scientists." One area of confusion is the tendency to equate equation imbalance with mixing orders of integration. Any time an analysis includes variables with different time series characteristics, we have mixed orders of integration. Consider, for example, an integrated variable, which contains a unit root and is nonstationary, where the mean, variance, and covariance vary over time, an example of which may be the gross domestic product (GDP). By contrast, a stationary series tends to hover around its mean over time, as seems to be the case for presidential approval in the USA, which exhibits autoregressive properties (Beck, 1991).[1] Some research concludes that presidential approval may be fractionally integrated (FI) (which would mean more persistence than an $I(0)$ stationary series) (Donovan *et al.*, 2020). In either case, given the different time series properties of presidential

---

[1]Stationary series are indicated as $I(0)$. Non-stationary series order $I(1)$, $I(2)$, or $I(3)$, depending on the number of times a series needs to be differenced to become stationary, although $I(1)$ is most common.

approval and GDP, an equation relating these variables mixes orders of integration. Importantly, this mixing does not necessarily produce an unbalanced equation.

Enns and Wlezien (2017) review the scholarly debate on the subject and clarify the distinction between imbalance and mixed orders of integration. Following Banerjee *et al.* (1993), they show that a linear combination of the regressand and of the regressors can produce a balanced equation, even if the variables are of different orders of integration. Using simulations and an applied example of models of income inequality in the United States, Enns and Wlezien (2017) demonstrate that there are situations where estimating general error correction/autoregressive distributed lag (GECM/ADL) models with different orders of integration does not produce an increased risk of spurious results. In this paper, we build on Enns and Wlezien's (2017) findings to further assess the estimation of GECM/ADL models when mixing orders of integration, specifically, to analyze the implications for the risk of type II error; that is, failing to reject a false null hypothesis. With good reason, researchers continue to be worried about type I error (e.g., Philips, this symposium; Kraft, Key, and Lebo, this symposium), and this concern motivated Enns and Wlezien's previous study on the subject. Yet, any recommended methods should be able to identify true relationships in the data when those exist, hence our focus in this paper.

We run simulations of a model with a stationary variable on the right-hand side and a dependent variable that contains both stationary and unit root, that is, integrated, components. As a result, we set up a data generation process in which the variables on the right- and left-hand sides are related but of different orders of integration. This setup is frequently found in social science research, where many variables may involve both stationary and unit root processes, such as people's attitudes over time (Converse, 1964; Achen, 1975; Erikson, 1979), party identification (Erikson *et al.*, 1998), vote intentions (Erikson and Wlezien, 2012), and global capital mobility (Ahlquist, 2006). Thus, focusing on this type of process allows us to demonstrate that mixing orders of integration does not necessarily result in an unbalanced equation, and holds implications for the myriad of published studies in the discipline that model these combined processes with GECMs.

Previous research has shown that the GECM/ADL avoids inflated type I error rates with this data-generating process (DGP). Our goal is to identify whether a GECM/ADL model can detect the true relationship underlying the generated data, and thus avoid type II error. Failure to do so would imply that the GECM/ADL is overly conservative when estimating relationships. By contrast, correctly identifying true relationships would suggest that the GECM/ADL can be appropriate in time series analysis with mixed orders of integration, at least where researchers can establish that equations are balanced (and other modeling assumptions are met).[2]

Our results suggest that the GECM/ADL can indeed identify the true relationship in our data. The mean coefficients also behave as expected, regardless of the autoregressive parameter (ρ) of the stationary component. This is most clear for the asymptotic case. In practice, when T is smaller, analyses are complicated by the fact that the observed time series characteristics of variables do not always match the underlying DGP. However, our simulations show that if researchers base their modeling decisions on the observed time series properties of the data (as is standard practice in time series analysis), they typically will reach sound inferences. Mixing orders of integration in a balanced equation evidently does not compromise the detection of true relationships between series. Our results underscore that pre-whitening data to equalize orders of integration is not always necessary, namely, in the presence of an already balanced equation. This is important, as Enns and Wlezien (2017) show that pre-whitening can come at the expense of identifying true relationships in the data.

This paper thus contributes to an increasing trend in the literature that asserts that there are exceptions to the idea that orders of integration have to be consistent across all series in a model

---

[2]Of course, balance is not the only criterion for effective time series modeling. Standard statistical assumptions apply, including independent and identically distributed (iid) residuals, i.e., "white noise" and, particularly with small samples, parsimony to avoid overfitting the model.

(Grant and Lebo, 2016, 4). Indeed, GECM/ADL models can be applied with mixed orders of integration, as long as equations are balanced and other assumptions are met as well.

1. Equation balance with mixed orders of integration does not necessarily increase the probability of type I error

Our starting point is the correct assertion that time series analysis requires equations to be balanced. Unbalanced equations suffer from mis-specification and typically result in an increased probability of type I error. Enns and Wlezien (2017) highlight that much of the emphasis in the discussion around equation balance has focused on an overly strict definition that entails having all variables belonging to the same order of integration. Banerjee *et al.* (1993, 164, italics ours) define an unbalanced equation as one "in which the regressand is not the same order of integration as the regressors, *or any linear combination of the regressors.*" Relying on Banerjee *et al.'s* (1993) definition, Enns and Wlezien (2017) show that because of the linear combination, regressors that contain different orders of integration can result in a balanced equation. This clarification is important, for many analyses in the social sciences are modeled with equations that mix orders of integration. Inappropriately restricting the concept of equation balance could thus lead to fruitless contestation of a multitude of published studies that mix orders of integration, but that actually have balanced equations that produce non-spurious results.

Two cointegrated I(1) series, when represented as a single equation GECM,[3] offer a classic case of a balanced equation with mixed orders of integration (Grant and Lebo, 2016; Keele *et al.*, 2016; Enns and Wlezien, 2017). The GECM is a balanced equation in this case because ΔY_t (the regressand) and ΔX_t are both stationary and the integrated regressors (X_{t-1} and Y_{t-1}) are jointly stationary.[4] Thus, both sides of the equation are stationary, which is necessary for proper estimation. Although illustrative, cointegration is not exceptional. In this symposium, Kraft, Key, and Lebo also highlight scenarios where linear combinations of integrated X or X and Y yield stationary processes, creating a balanced equation. Enns and Wlezien (2017) also demonstrate cases where it is appropriate to estimate models that mix orders of integration; specifically, as long as the equation is balanced (and others modeling assumptions are met), the GECM/ADL does not inflate the type I error rate.

Yet, some recent studies question the use of GECM/ADL based on type I errors (Grant and Lebo, 2016; Keele *et al.*, 2016), which seemingly contrasts with Enns and Wlezien's (2017) findings and other studies showing that when simulations are implemented correctly, the type I error rate follows the expected 5 percent (Enns *et al.*, 2016a, 2017b; Esarey, 2016). Like all methods, the GECM is only appropriate when relevant assumptions are met. Much of the recent confusion may stem from conducting simulations that violate these assumptions. Even when we specify the DGP in simulations, the observed series will not always reflect the underlying time series properties of the DGP because the simulated data also contain a stochastic component. If the simulations are designed to test a statistical approach that would be appropriate based on the DGP, but some of the simulated data departs from the specified DGP because of the stochastic component—which is especially likely when T is short—conclusions about the statistical approach may be misguided, because the statistical procedure is being evaluated on data it was not intended to be applied to. As Lebo and Grant (2016, 71) state, "Missteps here are easy if we diagnose the properties of our series in terms of some population instead of the sample in hand."

Indeed, the fact that small sample simulations do not necessarily reflect the time series properties specified in the DGP likely explains some (perhaps all) of the seeming divergent

---

[3]A GECM models the first difference of a variable in terms of the lagged level of that variable and the first difference and lagged level of each independent variable. Throughout this article, we refer to "GECM/ADL" because, although the GECM and ADL models look different, they are identical (algebraically equivalent) models.

[4]This is necessarily so if the two series are cointegrated.

conclusions across this symposium. For example, Kraft, Key, and Lebo in this symposium suggest that they observe inflated type I error rates because of equation imbalance. But equation imbalance cannot fully account for their results. To see why, consider their Figure 2, which shows that the rate of spurious regression declines as T increases. If the type I error rate in their simulations resulted because of equation imbalance in the DGP, we would not expect the false positive rate to decline as the sample size increases. To further evaluate the relationship between sample size and type I error rate in their simulations, we increased the sample size in their simulations to 5000 and then replicated their simulations designed to test the influence of adding unrelated $I(1)$ regressors. We chose $T = 5000$ so the data more closely follow the asymptotic time series properties of the DGP. Despite using the *exact same DGP*, increasing the sample size completely changed the conclusions of Kraft, Key, and Lebo's simulations; even with the inclusion of $I(1)$ regressors, the type I error rate drops to 5.2 percent with a mean value of −0.00002 (full results are reported in online Appendix 1).[5] Although adding unrelated regressors should never be advised in applied settings, doing so does not necessarily create an unbalanced equation.[6]

Since social scientists often confront small samples, it is of course critical to evaluate methods when T is small. However, when conducting these simulations, the series should be diagnosed and model assumptions tested before selecting the model (e.g., Enns *et al.*, 2016a, 2017b) or else both small and large samples should be simulated to assess whether the small sample properties deviate from the asymptotic results (e.g., Enns and Wlezien, 2017). Failure to follow these steps risks misinterpreting simulation results because the statistical model was chosen based on the asymptotic properties of the DGP instead of the properties of the data being analyzed.

Some of Philips' simulations in this symposium highlight a related concern. He evaluates three separate models (static, LDV (lagged dependent variable), and ARDL/GECM) on the same simulate data. In practice, theory and tests of statistical assumptions would almost always indicate one of these models was appropriate and the other two were inappropriate. What we typically want to know from data simulations is how a model performs when applied as it would be by a researcher. If the sample properties of the simulated data are ignored, and three models that would almost never be applied to the same data are evaluated, we would expect that at least two of the models would consistently perform poorly in the simulations. But the poor performance would be the result of evaluating the models when they should not have been applied (which a researcher could easily avoid with standard time series diagnostics).

In addition to not testing the time series properties of the simulated series to determine which model to estimate, it appears that Philips did not evaluate whether the coefficient on lagged X was significant prior to evaluating the Long Run Multiplier (LRM), which depends on that coefficient. Absent a significant relationship between X_{t-1} and ΔY_t in a GECM, a researcher has no reason to test for a significant LRM. And evaluating the LRM in this context will *necessarily* inflate the number of type I errors associated with this parameter. These are spurious associations that a practitioner would not encounter, because a practitioner should not estimate an LRM absent evidence of a long-term relationship.

These points offer three insights related to our paper. First, they are a reminder that particularly when T is short, simulations meant to guide applied research cannot be based on asymptotic properties of the DGP. We incorporate this insight into our small sample simulations later in this paper. Second, we believe this discussion helps account for the seemingly divergent findings across the symposium and between the symposium and past research. That is, what may at first look like different conclusions about equation balance and the rate of type I error rates with the GECM/ADL may disappear when simulations base modeling decisions on the observed

---

[5]This also follows mathematically. Kraft, Key, and Lebo's DGP creates a cointegrating relationship between y and $x1$ and the true weight on an irrelevant variable in the linear combination is zero, so an irrelevant variable should not create imbalance asymptotically.

[6]Also see Enns and Wlezien (2017).

data (as researchers do in practice). Finally, since Kraft, Key, and Lebo as well as Philips did not base their modeling decisions on the observed data, we should be careful drawing conclusions about published research based on their simulations.

2. Identifying true relationships with mixed orders of integration

As discussed, it has been demonstrated that unrelated stationary and first order integrated time series, that is, non-stationary series, can, in some cases, be analyzed together with a GECM/ADL model without concerns for spurious regressions (i.e., type I error). We now consider whether the GECM/ADL model can identify a true relationship between series that are of different orders of integration. Although time series researchers typically—and understandably—are more concerned about type I error, failing to detect true relationships in the data is also an issue. In the absence of knowledge of the effect of one variable on another, we would like to know that our estimation approach will reveal it.

To address this issue, we evaluate a regression model with a stationary variable on the right-hand side and a dependent variable that includes both stationary and unit root components. Wlezien (2000) refers to such a variable as a "combined" time series process. Here, the shock to a combined time series, e_t, can be separated into two parts: a series of stationary shocks that cumulate indefinitely (x_t^l) and another series x_t^S that decays (Wlezien, 2000, 79).[7] In theory, such series are integrated (Granger, 1980), as the portion that cumulates over time dominates.

There are many scenarios in the political and economic world that can produce combined time series. To begin with, consider that any process that includes long-term change and measurement error is such a series. But even putting aside measurement error, there are reasons to suppose that numerous processes combine both long-term and short-term change. Theories of people's attitudes over time reflect distinctions between effects that endure versus those that decay (Converse, 1964; Achen, 1975; Erikson, 1979). Characterizations of party identification also reflect these distinctions, and some scholars (Erikson *et al.*, 1998) explicitly conceive of macro-partisanship as a combined process. The same is true for electoral preferences, which clearly change over time, some of which lasts to impact the outcome and some does not (Erikson and Wlezien, 2012). We also see evidence of short-term and enduring changes in studies of political economy, such as the determinants of global capital mobility (Ahlquist, 2006) and its electoral consequences (Tomashevskiy, 2015). It may be that most seemingly "pure" integrated series actually are combined, where in addition to shocks to the series that cumulate over time there are shocks that decay. Indeed, any series that contains a unit root and is cointegrated with another series must combine integrated and stationary components. This can be seen in Kraft, Key, and Lebo's contribution to this symposium (Equations 5–7), where cointegrated Y combines X, which contains a unit root, and ζ, which is a stationary series. In sum, combined time series are common—and important—for political research, and notice they are substantially quite different to FI series (Box-Steffensmeier and Smith, 1998b), where all shocks decay, just more slowly than we expect of pure stationary processes.[8] Combined time series also are ideally suited for GECMs, which estimate both short run (stationary) and long run (integrated) components.

Just as important for our purposes, however, the data generation process of combined time series allow us to conduct simulations where the left- and right-hand side variables are related and of different orders of integration. Not only do researchers often find themselves analyzing

---

[7]A pure integrated series, by contrast, only contains a series of shocks that cumulate indefinitely. For more details, see Wlezien (2000).

[8]As Granger and Joyeux (1980b) showed, aggregating across AR processes with different rates of decay may produce an FI series. In practice, it may be difficult differentiating a combined and FI series, and differentiating each of these from a near-integrated series (De Boef and Granato, 1997), where shocks decay at a very slow rate (see Wlezien, 2000).

series with different orders of integration that are hypothesized to be related, but focusing on cases with different orders of integration offers an opportunity to further clarify the concept of equation balance. For our simulations,

$$Y_t = x_t^I + x_t^S \tag{1}$$

$$x_t^I = x_{t-1}^I + u1_t, \ u1_t \ N(0, 1) \tag{2}$$

$$x_t^S = \rho x_{t-1}^S + u2_t, \ u2_t \ N(0, 1) \tag{3}$$

where ρ equals 0.2, 0.5, or 0.8.

Because we want to evaluate whether the GECM/ADL can recover true relationships when the orders of integration on the right- and left-hand side of the equation are mixed, we estimate the equation,[9]

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_1 x_t^S + \beta_2 x_{t-1}^S + \delta \tag{4}$$

The equation does not include x_t^I, which means we are mixing a combined time series, Y, which in theory is integrated (Granger, 1980), with a stationary time series, x^S. Clearly, when analyzing combined time series (as with all data types), researchers should aim to model all explanatory factors. If x_t^I were correlated with x_t^S, this omission would create an omitted variable bias problem. Our goal, however, is to model related series that are of different orders of integration and omitting x_t^I ensures this scenario. Also keep in mind that time series researchers often face the possibility of omitted variables, either because repeated historical measures do not exist or because they simply do not know the true DGP. Consider research on electoral preferences mentioned earlier, where long-term effects of campaigns are difficult to directly capture on the right-hand side of models of pre-election polls (Erikson and Wlezien, 2012). Equation 4 would not be recommended if we knew the true DGP and had measures of all relevant variables, but it allows us to evaluate the performance of the GECM/ADL with mixed orders of integration and a common but imperfect specification. As a result, our simulations mirror the constraints that researchers may encounter in applied settings.

Since the true relationship between x_t^S and Y_t is 1.0, in our simulations we expect $\hat{\beta}_1$ to equal 1.0. (To be clear, the equation reveals the contribution the independent variable makes to our outcome variable, not the autoregressive parameter of the component.) Relatedly, we expect $\hat{\beta}_2$ to equal -1.0. That this is true can be seen by substituting for Y_{t-1}, which equals $x_{t-1}^I + x_{t-1}^S$, as follows:

$$Y_t = \alpha_0 + \alpha_1[x_{t-1}^I + x_{t-1}^S] + \beta_1 x_t^S + \beta_2 x_{t-1}^S. \tag{5}$$

Since Y contains a unit root, $\alpha_1 = 1$, and the equation reduces to

$$Y_t = \alpha_0 + x_{t-1}^I + x_{t-1}^S + \beta_1 x_t^S + \beta_2 x_{t-1}^S. \tag{6}$$

Given that $Y_t = x_t^I + x_t^S$, by construction, we expect β_2 to equal $-\beta_1$, which cancels the portion of x_{t-1}^S in Y_{t-1} (since x_{t-1}^S does not enter the DGP in Equation 1). Notice that this generalizes across

---

[9]Notice that this equation appears to be one of the cases that Kraft, Key, and Lebo identify where standard limiting distributions are appropriate, as x_t^S is stationary and there is little to no autocorrelation between the disturbances. Also note that while Equation 4 reflects the ADL specification, the GECM is the identical model, and as we show below, produces identical results.

combinations of x_t^I and x_t^S. For example, where the true relationship between x_t^S and Y_t equal to 2.0 (and that between x_t^I and Y_t equal to 1.0), we would expect β_1 in Equation 4 to equal 2 and β_2 to equal -2, which would again cancel out the portion of x_{t-1}^S in Y_{t-1}. Finally, we expect α_0 to equal 0.0. Of course, all of these theoretical expectations are based on the asymptotic case.

As previously noted, sometimes political scientists have suggested that it is never acceptable to mix orders of integration and that series should be pre-whitened to ensure the same order of integration for all variables in the model. However, despite different orders of integration on both sides of the equation, we show that the equation is indeed balanced. Substituting Equation 1 for Y, Equation 4 can be rewritten as follows:

$$x_t^S + x_t^I = x_{t-1}^S + x_{t-1}^I + \beta_1 x_t^S + \beta_2 x_{t-1}^S + \delta. \tag{7}$$

By rearranging Equation 7, in Equation 8, we see that we now have stationary series on both sides of the equation:

$$\Delta x_t^I + x_t^S = \beta_1 x_t^S + (1 + \beta_2) * x_{t-1}^S + \delta, \tag{8}$$

and so we would expect to be able to identify the true relationships between x^S and Y described above.[10] This does not mean that estimating Equation 4 will correctly represent the DGP, as it clearly does not, since we omit x_t^I to ensure that we mix orders of integration. We are intending only to illustrate that, just as an ADL (or GECM) does not necessarily induce spurious results when orders of integration are mixed, it can reveal true relationships with mixed orders of integration, at least when the equation is balanced (and other modeling assumptions are met). As De Boef and Keele (2008) have shown, the GECM/ADL does so in a general way, by allowing the data to determine the dynamic structure, that is, settling it empirically not by assumption.

2.1 The asymptotic case

We begin by presenting simulations where $T = 5000$, only to approximate the asymptotic behavior of the series, before turning to shorter, more realistic T. The asymptotic case is especially important because *if equation balance were a concern, it would be evident when T is large*. Table 1 presents the results of 2000 simulations where equations are estimated using the ADL, the GECM, and the differenced dependent variable (DV) specification.[11] Although the simulations employ a single lag, in practice, other lags could be considered and lag structure tested via goodness of fit statistics, such as the Akaike Information Criterion (see also, Hendry, 1995; Wilkins, 2018). The left part of the table presents the results for the ADL model. There we can see that the estimation almost always recovers the DGP described above. Specifically, the coefficient on x_t^s is consistently near 1.0 and the coefficients on x_{t-1}^s and Y_{t-1} are consistently equal and opposite signed.[12]

The GECM is mathematically equivalent to the ADL, so the results based on the former in the middle of Table 1 must be equivalent to the those for the latter (De Boef and Keele, 2008; Enns et al., 2016a). To see this, consider the ADL results in the first column, where α_1, the coefficient on $y_{t-1} = 0.9989$. If we subtract Y_{t-1} from both sides of the ADL, the dependent variable becomes ΔY, which matches the GECM and $\alpha_1^\star$, the coefficient on Y_{t-1} in the GECM, should equal $0.9989 - 1$, or -0.0011, which perfectly matches $\alpha_1^\star$ in the GECM results. The ADL coefficients

---

[10] Following Banerjee *et al.* (1993), no transformation is required to achieve balance in this case; what matters, "is the *possibility* of transforming" (168, italics in original). However, to achieve what Pickup (2020) call "*I*(0) balance," transformation is required.

[11] Table A.1 in online Appendix 2 reports analogous results where a disturbance term, q, is added to the DGP of Y_t.

[12] As noted above, asymptotically β_1 and β_2 will be equal and opposite, but as α_1 deviates from 1 in finite samples, this will be reflected in the estimate of β_2, as we see in Table 3 and online Appendix 3.

Table 1. Identifying a true relationship ($\beta_1 = 1.0$) between x_t^S and Y when x_t^S is stationary and Y combines stationary *and* unit root properties

	$T = 5000$																	
	ADL						GECM						First difference					
	$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%
$\hat{\alpha}_1$	0.9989	100	0.9989	100	0.9989	100	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
$\hat{\alpha}^*_1$	N/A	N/A	N/A	N/A	N/A	N/A	−0.0011	0	−0.0011	0	−0.0011	0	N/A	N/A	N/A	N/A	N/A	N/A
$\hat{\beta}_1$	0.9999	100	1.0000	100	1.0004	100	0.9999	100	1.0000	100	1.0004	100	0.9999	100	1.0000	100	1.0004	100
$\hat{\beta}_2$	−0.9983	100	−0.9994	100	−0.9995	100	N/A	N/A	N/A	N/A	N/A	N/A	−0.9994	100	−1.0005	100	−1.0006	100
$\hat{\beta}^*_2$	N/A	N/A	N/A	N/A	N/A	N/A	0.0016	5.6	0.0006	5.3	0.001	5.2	N/A	N/A	N/A	N/A	N/A	N/A

Notes: coef represents the mean coefficient estimate across 2000 simulations. % represents the percent of simulations for which we (correctly) reject the null hypothesis of no relationship. Consistent with expectations, across all models $\hat{\beta}_1 \approx 1.0$. As explained in the text, the other parameter estimates also follow expectations. ADL: $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_1 x_t^s + \beta_2 x_{t-1}^s + \delta$ GECM: $\Delta Y_t = \alpha_0 + \alpha^*_1 Y_{t-1} + \beta_1 \Delta x_t^s + \beta^*_2 x_{t-1}^s + \gamma$, where $\alpha^*_1 = \alpha - 1$ and $\beta^*_2 = \beta_1 + \beta_2$ from the ADL. First difference: $\Delta Y_t = \alpha^*_0 + \beta_1 x_t^s + \beta_2 x_{t-1}^s + \varepsilon$, where $\Delta Y_t = Y_t - \alpha_1 Y_{t-1}$ from the ADL.

on x_t^s and x_{t-1}^s (β_1 and β_2) equal 0.9999 and −0.9983, respectively. To show the equivalency with the GECM results, we can subtract and add $0.9999x_{t-1}$ to the right hand side of the ADL results. This produces $0.9999x_t^s - 0.9983x_{t-1}^s - 0.9999x_{t-1}^s + 0.9999x_{t-1}$, which equals $0.9999\Delta x + 0.0016x_{t-1}$, exactly the same as the estimates for β_1 and $\beta_2^\star$ in the GECM results. These results illustrate that the ADL and GECM *always* produce the *exact* same numerical information; the only difference is how the two models present this information.

The right side of Table 1 presents the results from a model that first-differences the dependent variable. The model offers an additional way to show that the previously estimated ADL/GECM model is balanced even though the model mixes different orders of integration. To see why, recall that the right hand side of the model (x_t^s) is stationary and the dependent variable contains a unit root (as well as stationary properties). The standard approach in such a situation is to first-difference Y to make it stationary. Because $\Delta Y = Y_t - Y_{t-1}$, as the series approaches infinity, subtracting Y_{t-1} from both sides of the ADL in Equation 4 yields the first-difference model. In other words, given our sample size and DGP, the first-difference model the first-difference model approximates the ADL (and GECM). The results on the far right of Table 1 support this prediction. In most cases the numerical results are identical. The largest difference is 0.001. As can be seen in online Appendix 3, things are nearly identical when $T = 200$, 100, and 50, although we consider these simulations in more detail below. In other words, in cases where the data could suggest either the GECM/ADL or the first-difference model, both yield nearly identical results.[13]

As discussed above, Equation 4 is misspecified by construction, because it omits x_t^I.[14] Clearly, theory must guide model specification and it would be wrong to conclude that these results imply that the GECM/ADL is always appropriate. They key point is that even with mixed orders of integration, we find that the estimated ADL/GECM in Equation 4 is balanced. Our initial focus, however, has been on an atypically large sample size ($T = 5000$), so the observed data follow the asymptotic properties.

2.2 Diagnosing and estimating in practice

In practice, when dealing with finite time series where the true characteristics of the variables and their relationships are unknown, before selecting a model, researchers would first identify the characteristics of the variables. For instance, scholars commonly employ tests, such as the augmented Dickey–Fuller (ADF) test, to diagnose whether series are nonstationary and whether there is drift and/or trend as well.[15]

To illustrate this process, we return to the series we generated for Table A.2 in online Appendix 3. As before, Y always combines unit root and stationary properties and x_t^s is always stationary, where the autoregressive parameter $\rho = 0.2$, 0.5, or 0.8 and $T = 50$, 100, or 200. Instead of analyzing the data naively, we diagnose the time series properties on each of the simulated combined DVs as well as the stationary independent variables (IVs) with ADF tests, which

---

[13]Specifically, with these sample sizes, the ADL almost always yields accurate inferences about the presence of a relationship (between 99.2 and 100 percent of the time), $\hat{\beta}_1$ is always within 0.01, and the estimates for α_1 and x_t^s are offsetting (within 0.01), as expected.

[14]Thus, any correlation between $u1_t$ and $u2_t$ in Equations 2 and 3 would bias the estimate of β_1.

[15]The ADF tests for the presence of a unit root in a variable, where the first difference of the variable is regressed on its lagged level and a series of lagged differences, as follows:

$$\Delta Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 \Delta Y_{t-1} + \beta_3 \Delta Y_{t-2} + \cdots + \Delta Y_{t-x}. \tag{9}$$

The coefficient for the lagged level variable is the primary focus: if it is indistinguishable from 0, then we conclude that the variable is integrated; if it is negative and significantly different from 0, we conclude that the variable is stationary. Note that the appropriate critical values for unit root test are nonstandard. For a more complete discussion and well as an introduction to other tests scholars use, see e.g., Pickup (2014); Box-Steffensmeier *et al.* (2014). Particularly with small samples, some series may not appear to be strictly $I(1)$ or $I(0)$, but FI instead (e.g., Baille, 1996; Lebo and Clarke, 2000).

Table 2. Percent of simulations that identify Y and x_t^s as $I(1)$ or $I(0)$ with ADF tests when $T = 100$ and $\rho = 0.5$

Y $I(1)$, x_t^s $I(0)$	78.3
Y $I(0)$, x_t^s $I(0)$	21.4
Y $I(1)$, x_t^s $I(1)$	0.3
Y $I(0)$, x_t^s $I(1)$	0.1

Notes: The first row shows the percent of simulations that correctly identify both series.

are particularly relevant for combined time series (see Wlezien, 2000). To reiterate, while we know the characteristics of the underlying process, researchers in the observational world will not know, so it is important for us to mimic that process, as it has implications for the estimation approaches that are appropriate. As Durr (1992, 193) explains, "empirical diagnoses of time-series data are necessarily a function of a finite sample of a realization of the process in question" (also see Enns *et al.*, 2016a; Lebo and Grant, 2016: 71–72).

To illustrate, let us take the case where the ρ of the stationary component is 0.5 and $T = 100$. We apply ADF tests using a critical value of 0.05 to each of the 2000 simulated series of both our DV and IV. Table 2 summarizes the results. The first two rows show that we always correctly identify that the IV (x_t^s) is stationary but have more difficulty with the combined DV, and identify the underlying integrated process only 78.6 percent of the time (which we see by summing the first and third rows). This pattern comports with previous research (Wlezien, 2000). By implication, we correctly infer both underlying processes in 78.3 percent of the simulations. Much less frequently—21.4 percent of the time—we conclude that both series are stationary.

These inferences are important, as they matter for estimation. For instance, in the case where both variables appear stationary, we can proceed to estimate an ADL model. Where the DV is nonstationary and the IV stationary, time series practice would recommend regressing the differenced DV on the current and lagged IV. The results of estimating these models are summarized in Table 3. Here we can see that in the 78.3 percent of the simulations where we correctly identify the underlying true processes and estimate a model with a differenced DV, we detect a significant effect of the IV in each of our 2000 simulations, with a mean coefficient of near-perfect 1.0 (0.9999). We also detect the expected β_2 (i.e., x_{t-1}^s) in every simulation, the mean estimate of which is -1.0031. The virtually equal and oppositely-signed effects of β_1 and β_2, which correspond with x^s and x_{t-1}^s, imply that the difference in our DV, Y, reflects the difference in x_t^s, that is, $\Delta Y = \Delta x^s$.[16] In 21.4 percent of the simulations where both variables appear stationary and we estimate an ADL (as a researcher would do in an applied setting), we obtain similar results, identifying significant current and lagged effects 100 percent of the time and with mean coefficients of 1.0001 and -0.9491, respectively. The estimates do vary, of course, which is important; we nevertheless always detect a true effect of our IV on our DV.

Let us now consider different combinations of ρ and T, where the former varies between 0.2, 0.5, and 0.8 and the latter between 50, 100, and 200. Although we know the DGP, we are interested in how researchers would diagnose these series in an applied setting. Table 4 shows the results of our diagnostic analyses for each set of simulations. The patterns are similar to what we saw in Table 2 (and 3), excepting where ρ is large (0.8) and T is low (50). In this case, we frequently identify integration in both the DV and the IV, as the (low) power of the ADF tests make it difficult to reject the null of nonstationarity.[17] This is important because it implies a different estimate strategy. Specifically, with two $I(1)$ variables, we need to first assess whether they

---

[16]That is, we know that $\beta_1 x_t^s + \beta_2 x_{t-1}^s = \beta_1 x_t^s - \beta_1 x_{t-1}^s + \beta_2 x_{t-1}^s + \beta_1 x_{t-1}^s = \beta_1 \Delta x_t^s + \beta_2 x_{t-1}^s + \beta_1 x_{t-1}^s$. Thus, multiplying through using the mean coefficients in Table 3 we get $0.9999 \Delta x_t^s - 0.003 x_{t-1}^s$. Again, this comports with the results from estimating the GECM in Table 1.

[17]The high rate of acceptance (66.7 percent) of the null of nonstationarity for our IV reflects the combination of a relatively large ρ and small T. But even with these parameters, the evidence in favor of nonstationarity is not strong, for example, we would reject the null of a unit root in all but 6.75 percent of cases using a critical value of 0.10 instead of 0.05.

TABLE 3. Results from simulated first difference and ADL models with different combinations of independent and dependent variables

| | First difference, $Y = I(1)$, $x_t^s = I(0)$ | | ADL, $Y = I(0)$, $x_t^s = I(0)$ | |
	coef	%	coef	%
$\hat{\alpha}_1$	N/A	N/A	0.9461	100
$\hat{\beta}_1$	0.9999	100	1.0001	100
$\hat{\beta}_2$	−1.0031	100	−0.9491	100
% cases		78.3		21.4

Notes: Consistent with expectations, across all models $\hat{\beta}_1 \approx 1.0$. As explained in the text, the other parameter estimates also follow expectations. The bottom row (%) shows the percent of simulations that correctly identify the series as $Y = I(1)$, $x_t^s = I(0)$ or $Y = I(0)$, $x_t^s = I(0)$. First difference: $\Delta Y_t = \alpha^*_0 + \beta_1 x_t^s + \beta_2 x_{t-1}^s + \varepsilon$ ADL: $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_1 x_t^s + \beta_2 x_{t-1}^s + \delta$

Table 4. Percent of simulations that identify Y and X as $I(1)$ or $I(0)$ with ADF tests when $T = 50$, 100, and 200

| | $T = 50$ | | | $T = 100$ | | | $T = 200$ | | |
	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
Y $I(1)$, X $I(0)$	75.6	79.9	29.7	73.6	78.3	73.6	76.4	77.4	82.8
Y $I(0)$, X $I(0)$	24	16.5	3.7	26.3	21.4	11.7	23.6	22.6	17.1
Y $I(1)$, X $I(1)$	0.5	3.2	61.7	0.2	0.3	13.1	0.1	0.1	0.2
Y $I(0)$, X $I(1)$	0	0.5	5.0	0	0.1	1.8	0	0	0

Notes: The first row shows the percent of simulations that correctly identify Y as $I(1)$ and X as $I(0)$. ρ indicates the value of the autoregressive parameter in X. Results based on ADF tests on each of the 2000 simulated DVs and IVs using the particular ρ and T, as indicated. Columns may not sum to 100 due to rounding.

are cointegrated. If they are, we should proceed to a general model suitable for cointegrated series like the GECM; if not, it is necessary to difference both the DV and IV and to estimate a more restricted model.[18] This is standard time series practice. Furthermore, when we estimate the first-difference models that meet these criteria, we identify the expected relationship in all situations.

In sum, two patterns stand out in these results. First, even in the case where ρ is large (0.8) and T is low (50), most of the time the ADF identifies the time series properties of the DGP. Second, when the ADF indicates the finite series differs from the time series properties of the DGP (which becomes more likely with smaller samples), if we estimate a model based on the observed properties of the data (as indicated by the ADF), we still accurately identify the true relationship between the series in most cases.

3. Conclusion

Time series researchers have understandably focused their studies on the issue of how to avoid spurious correlations. In this paper, we turned our attention to the capacity of GECM/ADL models to detect true relationships between series of different orders of integration, as long as equations are balanced (and other modeling assumptions are met). Although it is known that the GECM/ADL avoids spurious regression in this situation (Enns and Wlezien, 2017), if the model cannot detect true relationships in the data, results would not be informative to researchers.

Our simulations show that an equation with a stationary variable on the right-hand side and a dependent variable that combines both unit root and stationary components (a "combined" time series process) can be estimated using a GECM/ADL model. In the simulations, we are able to identify the true relationship between the variables in the asymptotic case. This means that mixing orders of integration not only does not necessarily increase the probability of type I error, as shown by Enns and Wlezien (2017), it also does not necessarily increase the risk of type II error.

---

[18]See De Boef and Keele (2008) and Keele *et al.* (2016) for more on the value of moving from a general to a specific model.

Of course, things are more complicated when dealing in finite time. As we show in the paper, researchers must first identify the time series properties of the observed data; after doing so, assuming other modeling assumptions are met, the GECM/ADL consistently identifies true relationships.

These findings have important implications for research in most areas in political science. First, they underscore that equation balance and mixed orders of integration, while related, are not one in the same, and that it is possible to have both. Second, we thus need not avoid estimation with mixed orders of integration, or rule out previous research based on such estimation, at least where we have equation balance. Third, and more specifically, with a balanced equation, we can detect true relationships (and eschew spurious ones) even in cases where we cannot correctly represent the DGP. This is of special importance because we often are uncertain about the characteristics of our variables. Although the necessity of pre-testing is encouraged in the literature, including in this symposium (Philips), detecting the properties of series is a difficult task. The findings from our simulations thus assuage some of the problems practitioners often face while also highlighting the importance of the general approach that DeBoef and Keele (2008) recommend. Of course, even when a parsimonious model guided by complete pre-tests fits the data well, standard diagnostics must be evaluated, particularly the characteristics of the residuals. This is (or at least should be) routine in time series analysis, as it provides critical information to researchers.

Finally, we make the additional point that, while there is consensus that the GECM avoids spurious regression when assumptions are met and it is implemented correctly (e.g., Enns et al., 2016a, 2017b; Grant and Lebo, 2016), this symposium highlights that there is still some disagreement about how frequently these assumptions are met. We understand that part of the disagreement stems from the application of different approaches to simulations. In our minds, a focus on both large and small samples for simulations, estimating models on simulated data based on the observed data (not naively based on the DGP), and attention to both type I and type II error rates will help advance our understanding of the limits of GECM/ADL models and the use of simulations in time series analysis.

References

Achen C (1975) Mass political attitudes and the survey response. *American Political Science Review* **69**, 1218–1231.

Ahlquist JS (2006) Economic policy, institutions, and capital flows: portfolio and direct investment flows in developing countries. *International Studies Quarterly* **50**, 681–704.

Baille RT (1996) Long memory processes and fractional integration in econometrics. *Journal of Econometrics* **73**, 5–59.

Banerjee A, Dolado J, Galbraith JW and Hendry DF (1993) *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Beck N (1991) Model Selection: Are Time Series Techniques Useful in Cross-Sectional Problems? 999: Midwest Meeting.

Box-Steffensmeier JM, Freeman JR, Hitt MP and Pevehouse JCW (2014) *Time Series Analysis for the Social Sciences*. New York: Cambridge University Press.

Box-Steffensmeier JM and Smith RM (1998) Investigating political dynamics using fractional integration methods. *American Journal of Political Science* **42**, 661–689.

Converse PE (1964) The Nature of Belief Systems in Mass Publics. In Apter DE (ed.), *Ideology and Discontent*. Ann Arbor: University of Michigan Press, pp. 206–261.

De Boef S and Granato J (1997) Near-integrated data and the analysis of political relationships. *American Journal of Political Science* **41**, 619–640.

De Boef S and Keele L (2008) Taking time seriously. *American Journal of Political Science* **52**, 184–200.

Donovan K, Kellstedt PM, Key EM and Lebo MJ (2020) Motivated reasoning, public opinion, and presidential approval. *Political Behavior* **42**, 1201–1221.

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

Enns PK and Wlezien C (2017) Understanding equation balance in time series regression. *The Political Methodologist* **24**, 2–12.

Enns PK, Kelly NJ, Masaki T and Wohlfarth PC (2016) Don't Jettison the general error correction model just yet: a practical guide to avoiding spurious regression with the GECM. *Research and Politics* **3**, 1–13.

Enns PK, Kelly NJ, Masaki T and Wohlfarth PC (2017) Moving forward with time series analysis. *Research and Politics* **4**, 1–7.

Erikson RS (1979) The SRC panel data and mass political attitudes. *British Journal of Political Science* **9**, 89–114.

Erikson RS and Wlezien C (2012) *The Timeline of Presidential Elections*. Chicago: University of Chicago Press.

Erikson RS, MacKuen MB and Stimson JA (1998) What moves macropartisanship? A reply to Green, Palmquist, and Schickler. *American Political Science Review* **92**, 901–912.

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

Freeman JR (2016) Progress in the study of nonstationary political time series: a comment. *Political Analysis* **24**, 50–58.

Granger CWJ (1980) Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* **14**, 227–238.

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

Grant T and Lebo MJ (2016) Error correction methods with political time series. *Political Analysis* **24**, 3–30.

Hendry DF (1995) *Dynamic Econometrics*. Oxford: Oxford University Press.

Keele L, Linn S and McLaughlin Webb C (2016) Treating time with all due seriousness. *Political Analysis* **24**, 31–41.

Lebo MJ and Clarke HD (2000) Modelling memory and volatility: recent advances in the analysis of political time series. *Electoral Studies* **19**, 1–7.

Pickup M (2014) *Introduction to Time Series Analysis*. Thousand Oaks, CA: Sage.

Pickup M and Kellstedt P (2020) Equation Balance in Time Series Analysis: What it is and How to Apply It. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3526534.

Tomashevskiy A (2015) Capital preferences: international capital and government partisanship. *International Studies Quarterly* **59**, 776–789.

Wilkins AS (2018) To lag or not to lag?: Re-evaluating the use of lagged dependent variables in regression analysis. *Political Science Research and Methods* **6**, 393–411.

Wlezien C (2000) An essay on 'Combined' time series processes. *Electoral Studies* **19**, 77–93.