

RESEARCH NOTE

Equation balance in time series analysis: lessons learned and lessons needed

Mark Pickup 

Department of Political Science, Simon Fraser University, Burnaby, Canada
Corresponding author. Email: mark.pickup@sfu.ca

(Received 21 May 2021; revised 9 September 2022; accepted 13 September 2022; first published online 11 October 2022)

Abstract

The papers in this symposium use Monte Carlo simulations to demonstrate the consequences of estimating time series models with variables that are of different orders of integration. In this summary, I do the following: very briefly outline what we learn from the papers; identify an apparent contradiction that might increase, rather than decrease, confusion around the concept of a balanced time series model; suggest a resolution; and identify a few areas of research that could further increase our understanding of how variables with different dynamics might be combined. In doing these things, I suggest there is still a lack of clarity around how a research practitioner demonstrates balance, and demonstrates what Pickup and Kellstedt (2021) call $I(0)$ balance.

Keywords: Autoregressive distributed lag model; equation balance; general error correction model; time series

1. Lessons learned and potential confusion

This symposium contains three papers that use Monte Carlo (MC) simulations to demonstrate the consequences of estimating time series models with variables that are of different orders of integration. In each paper, the concept of *balance*¹ is implicitly or explicitly raised as an important consideration in understanding the consequences of mixing orders of integration. While balance has been discussed to a moderate degree in the econometrics literature, it has had relatively little impact on political science. Very little empirical work in political science that uses longitudinal analysis raises the issue of balance to defend the theoretical or empirical model used. The discussion of balance has been mostly relegated to methodological discussions amongst a small group of academics (Freeman, 2016; Keele *et al.*, 2016a,b; Lebo and Grant, 2016; Enns and Wlezien, 2017), many of whom are involved in this symposium.

One reason the concept of balance has not translated well into empirical work may be the lack of clarity about its meaning. While there seems to be agreement on the definition of balance, there is a lack of clarity about how model balance is determined and the exact consequences of having a model that is not balanced. The articles in this symposium go some way to providing this clarity, at least for a limited range of models. Each paper looks at a different combination of variables and asks different questions about the consequences but all of them focus on the Autoregressive Distributed Lag (ARDL) model and/or the isomorphic General Error Correction Model (GECM).² In this summary, I very briefly outline what we learn (and don't

¹In this symposium, a model is defined as balanced “if and only if the regressand [left-hand side] and the regressors [right-hand side] (either individually or collectively, as a co-integrated set) are of the same order of integration” (Banerjee *et al.*, 1993, page 166).

²Isomorphic in the sense that they have the same properties with respect to their estimation of relationships between variables.

learn) from the papers; identify an apparent contradiction that might increase, rather than decrease, confusion around the concept of balance; suggest a resolution; and identify a few areas of research that could further increase our understanding of how variables with different dynamics might be combined. Before doing so, I would like to thank the authors of the three symposium papers for their contributions to the field of longitudinal data analysis in political science. These papers have challenged me to think more carefully about combining variables of different orders of integration and I am certain it will do the same for others. I agree with most of what has been written. The disagreements that I do discuss below are intended as good-natured, with the intent of providing further clarity to the research practitioner that would like to understand how the concept of balance applies to their empirical work.

Starting with Kraft, Key, and Lebo (KK&L), they examine the GECM containing multiple (>2) $I(1)$ variables in the cointegrating equation to demonstrate the consequences of what they perceive to be a common misinterpretation of the error correction coefficient. As an example, the GECM with y_t , $x_{1,t}$, and $x_{2,t}$ in the cointegrating equation is:

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 \Delta x_{1,t} + \beta_2 x_{1,t-1} + \beta_3 \Delta x_{2,t} + \beta_4 x_{2,t-1} + \epsilon_t \quad (1)$$

KK&L argue that it is common to interpret the null and alternative hypotheses tested by the error correction coefficient (α_1) and its standard error as:

- H_0 : y is not co-integrated with all x ;
- H_A : y is co-integrated with all x .

They argue that the correct null and alternative hypotheses are:

- H_0 : y is not co-integrated with any x ;
- H_A : y is co-integrated with at least one x .

The consequence of this misinterpretation is that practitioners are using the rejection of the null hypothesis as evidence that there is co-integration between all the variables in the empirical model when, in fact, the empirical model includes variables that are not in the data generating processes (DGP). In other words, practitioners are concluding that there are co-integrating relationships when there are not. The connection to balance is that equation (1) is $I(0)$ on both sides and balanced if y_t is co-integrated with all x_t . Otherwise, it is $I(0)$ on the left-hand side, $I(1)$ on the right-hand side and not balanced. If practitioners are making the error that KK&L claim, they are treating the empirical model as balanced when this is not necessarily the case.

KK&L demonstrate the consequences through MC simulations using empirical models that include $I(1)$ variables in the co-integrating equation that are not in the DGP (and therefore, not co-integrated with y_t). Most problematic is the false detection of relationships between y_t and unrelated x_t (Type 1 errors). This demonstrates how the inclusion of a variable in the empirical model that is not in the DGP can unbalance an otherwise balanced model. This is a valuable lesson. Researchers have long been warned about the dangers of including extraneous variables in their empirical models without carefully thinking about how they might affect the estimates of the key parameters of interest (e.g., Achen, 2005). This is a time series specific version of the same warning. Fortunately, for research practitioners, there are other procedures for determining the co-integrating relationships between multiple variables (e.g., Johansen, 1988; Stock and Watson, 1993; Pesaran *et al.*, 2001; Philips, 2018; Webb *et al.*, 2019). Unsurprisingly, these procedures require the use of more than one test statistic to determine the cointegrating relationship between y_t and multiple x_t . I am not entirely convinced many practitioners really believe that the single test statistic based on the estimate of α_1 is sufficient but, to the extent that practitioners do make the error highlighted by KK&L, the demonstration of the consequences is a useful one.

KK&L also raise other issues with the use of GECMs in political science, which they do not pursue in their contribution to the symposium. The most important are that the possibility of cointegration between more than two variables opens up the possibility of: (1) multiple cointegrating vectors; and (2) more than one equation (i.e., more than one endogenous variable). As KK&L note, most political scientists assume a single cointegrating vector and a single equation model will suffice. This is about all they say on the subject, which is notable given that it is highly unlikely that these assumptions are valid in most applications.

Finally, the MC simulations conducted by KK&L should probably have been extended in one way. They examine the consequences of including $I(1)$ variables in the empirical model that are not in the DGP. It would have been useful to test the consequences of including $I(1)$ variables in the empirical model that are not part of the cointegrating equation but are in the DGP through their first difference. In other words, the variables have a short-run affect on y but no long-run equilibrium relationship with y . Of course, there are always more simulations that could be run and KK&L had to make choices but this is a simple extension that research practitioners are likely to run into. It would be helpful for them to know how likely they are to falsely reject the null of no long-run relationship under these circumstances.

Philips examines the probability of a Type 1 error for a variety of time series DGPs and empirical models. In terms of empirical models, he looks at the static and lagged dependent variable (LDV) models, in addition to the ARDL and GECM. In terms of the DGPs, he includes (1) $y_t \sim I(0)$ and an unrelated $x_t \sim I(0)$, (2) $y_t \sim I(0)$ and an unrelated $x_t \sim I(1)$, (3) $y_t \sim I(1)$ and an unrelated $x_t \sim I(0)$, and (4) $y_t \sim I(1)$ and an unrelated $x_t \sim I(1)$. Philips uses MC simulations to determine the probability of a Type 1 error and comes to a number of different conclusions. One of which is the importance of dynamic completeness, as described by Wooldridge (2010). Essentially, it is important to include (at least) as many lags of the independent and dependent variables in the empirical model as exist in the DGP. Other conclusions relate to the problems of mixing certain orders of integration in an empirical model, and of regressing an $I(1)$ variable on another $I(1)$ variable when they do not cointegrate.

Philips claims to be agnostic about the issue of balance but implicitly he is relying on balance to develop recommendations. A key recommendation made by Philips is pretesting the variables to be used in the empirical model to determine orders of integration and co-integration. Assuming the researcher is able to produce reliable pretesting results, he provides recommendations about when and how to change the empirical model (through the judicious first differencing of variables) in order to reduce Type 1 errors. Even though Philips does not make it explicit, these recommendations are based on his understanding of how the model needs to be changed to produce a balanced equation, or what (Pickup and Kellstedt, 2021) call $I(0)$ balance.

Phillips only discusses and tests a limited range of scenarios (six in total). This is sufficient to make the point that mixing some orders of integration creates problems of inference. However, because he does not explain how he is using balance to determine the correct model to use, the reader does not learn how to generalize the results beyond the six scenarios. This is a missed opportunity. Also because Phillips does not discuss balance, it is also not apparent to the reader that of the scenarios that produce problematic estimates, two of them are violations of balance and one of them is a violation of $I(0)$ balance. Because this is not apparent to the reader, it is also not apparent that not all violations of balance may be equally problematic. This is important and even if Phillips did not have the space to explore why this was the case, he could have noted it.

Phillips' recommendations are also made problematic by the difficulties of producing reliable pretesting results. The literature is filled with examples of how unreliable these tests can be when T is small (or even moderately large), as it is in many empirical applications (e.g., Cochrane, 1991). That said, the Philips paper does highlight the problems of mixing some orders of integration in certain models, and makes a case for doing more than is the norm in political science to think carefully about the orders of integration of the variables being used in an empirical model. This alone, is a useful contribution.

Enns, Moehlecke, and Wlezien (EM&W) also use MC simulations. They do so to look at the consequences of using an ARDL or the equivalent GECM when y_t is first-order integrated and x_t is stationary.³ EM&W use simulations to demonstrate that an ARDL/GECM empirical model can successfully estimate the relationship between a right-hand side variable that is $I(0)$ stationary and a dependent variable that is $I(1)$ (with an autoregressive error) (EM&W, page 19), as long as the empirical model is balanced and T is moderately large. They claim these results also apply when T is not large. In this symposium they focus on showing this does not increase the rate of Type 2 errors but they use a similar set of simulations in their previous work to make the case that these conditions also do not increase the risk of Type 1 errors (Enns and Wlezien, 2017). That practitioners can reliably estimate a ARDL/GECM with first order integrated y_t and a stationary x_t appears to run counter to Philips. Philips concludes that spurious findings (Type 1 errors) are increased when the dependent variable is unit root and the independent variable is stationary and that the practitioner should first difference y_t before including it in the empirical model.⁴ This is a potential source of confusion to the symposium reader.

EM&W offer an explanation for why their results differ from those of Philips. The argument is that the series used in simulations do not always reflect the properties of the intended DGP because the simulated data contain a stochastic component. They argue this is especially likely when T is small (e.g., 50). They argue that this is demonstrated by the fact that the performance of the GECM improves once T is large (e.g., 5000). EM&W use the same line of reasoning to explain why they disagree with the results of KK&L.

There are three reasons why I don't think this explains the differences between Philips and EM&W (or KK&L and EM&W). First, the stochastic component is part of the DGP. It is not, as EM&W seem to imply, the difference between the DGP and the simulated data. Each simulated data set is one realization of the stochastic DGP and the fact that it deviates from the deterministic component of the DGP is a feature and not a flaw. Second, when reporting mean coefficient estimates from the simulations (as is done in this symposium), the question is not whether individual simulated data sets deviate from the deterministic component of the DGP. The question is whether the simulated data sets reflect the deterministic component on average. By the law of large numbers, the mean coefficient estimate from the simulated data sets will converge on the DGP value as the number of simulated data set increases—not as T increases, as EM&W seem to suggest. Third, the magnitude of T (and the variance of the stochastic component) will affect the variance of the estimated coefficients from the simulations but again this is a feature and not a flaw. The specification of T is part of the DGP. Increasing T changes the DGP. The variance (but not the mean) may decrease by increasing T but that is the actual variance of the estimator for the specified DGP, not a discrepancy between the DGP and the simulated data. EM&W's demonstration that the estimates change when T is increased is a demonstration that the model (and estimator) performs differently when the DGP is different.

This all is not to say that a change in the performance of a model/estimator as the DGP changes (through an increase in T) is not interesting and potentially important. And the question remains why the performance of the ARDL/GECM with a first order integrated y_t and a stationary x_t is problematic when the T of the DGP is small (e.g., 50) but improves when the T of the DGP is increased. I attempt to address this puzzle in the next section, and in doing so I hope to resolve the confusion that may be created by the apparently contradicting results within the symposium.

Before doing so though, it is important to note that EM&W also suggest that the simulations run in the other two symposium papers don't take into account how practitioners actually work.

³When defining the y_t variable, EM&W draw a distinction between an $I(1)$ variable and a variable that combines $I(1)$ and $I(0)$ processes. However, all $I(1)$ processes contain an $I(0)$ process. In $x_t = x_{t-1} + \epsilon_t$, ϵ_t is $I(0)$. There is no distinction to be made here. The distinction that EM&W actually seem to making is between an $I(1)$ series with a static error component (ϵ_t) and an $I(1)$ series with a first-order autoregressive error component. They call this a "combined process" but as they note, it is an $I(1)$ process.

⁴A spurious regression problem occurs when the t -statistics for the slope parameters indicate a relationship much more often than they should at the chosen test level.

They note that when practitioners estimate models, they rarely do so without first testing the dynamic properties of the variables to be included in the model. Therefore, they argue, simulations that don't take this into account are not fairly evaluating the performance of the models/estimators being tested. I think it is entirely fair to say that many practitioners employ pre-estimation testing and this can influence which model/estimator the practitioner uses.⁵ Therefore, the results of simulations that do not take these pre-estimation procedures into account must be interpreted accordingly. They tell us what happens if a practitioner blindly applies a model/estimator. The use of pre-estimation testing could improve the outcome or it could make it worse, depending on the pre-estimation procedure used. As I noted earlier, pre-estimation procedures are problematic and can lead to incorrect conclusions regarding orders of integration and/or the presence/absence of cointegration. In fairness to Philips and KK&L, they do not entirely ignore the potential use of pre-estimation procedures. Their recommendations include this type of testing in order to avoid the problems identified by their simulations.

2. Resolving the confusion?

To understand why Philips and EM&W reach somewhat different conclusions, we first look at the basis of the argument in EM&W. They use the following DGP in their MC simulations:

$$y_t = x_t^I + x_t^S \tag{2}$$

$$x_t^I = x_{t-1}^I + \mu_{1,t} \tag{3}$$

$$x_t^S = \rho x_{t-1}^S + \mu_{2,t} \tag{4}$$

where $\mu_{1,t}$ and $\mu_{2,t}$ are NID(0,1). It is the relationship between y_t and x_t^S that they claim can be estimated using an ARDL/GECM. Given the interest in this particular relationship, it should be noted that the DGP can be rewritten as follows:

$$\begin{aligned} y_t &= x_t^I + x_t^S \\ y_t &= x_{t-1}^I + \mu_{1,t} + x_t^S \\ y_t &= x_{t-1}^I + x_{t-1}^S + x_t^S - x_{t-1}^S + \mu_{1,t} \\ y_t &= y_{t-1} + \Delta x_t^S + \mu_{1,t} \\ \Delta y_t &= \Delta x_t^S + \mu_{1,t} \end{aligned} \tag{5}$$

This represents a static relationship between two stationary variables (Δy_t and Δx_t^S) and the

⁵One aspect of this criticism in EM&W does not seem accurate. Philips examines how well different models estimate short- and long-run effects. EM&W argues that a practitioner would/should not estimate the long-run effect of x_t if the coefficient on x_{t-1} is not statistically significant in the GECM, and so Philips' performance statistics for the long-run effect are inaccurate. I disagree that the lack of statistical significance for the coefficient on x_{t-1} means that a practitioner should not estimate a long-run effect. Presumably, EM&W's logic is that the coefficient on x_{t-1} is the numerator for the long-run effect estimate, and so if this coefficient is equal to 0, so is the long-run effect. However, a failure to reject the null hypothesis that the coefficient on x_{t-1} is 0 does not necessarily mean it is 0. We may simply not have the power to reject the null. Meanwhile, the long-run effect (if it is non-zero) will be larger than the coefficient on x_{t-1} and we often have more power to detect it.

appropriate empirical model would be:

$$\Delta y_t = \beta \Delta x_t^S + \epsilon_t \tag{6}$$

which should provide an unbiased estimate of $\beta = 1$. This is the short-run effect of x_t^S on y_t . This empirical model correctly assumes that x_t^S has no long-run effect on y_t .⁶ However, EM&W would like to demonstrate that the ARDL/GECM can provide an unbiased estimate of this relationship, so they use the following ARDL empirical model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t^S + \beta_2 x_{t-1}^S + \epsilon_t \tag{7}$$

EM&W argue that (7) is balanced and they do so in the following way. They note the DGP for Y is: $y_t = x_t^I + x_t^S$. They suggest this can be substituted into equation (7):

$$x_t^I + x_t^S = \alpha_0 + \alpha_1(x_{t-1}^I + x_{t-1}^S) + \beta_1 x_t^S + \beta_2 x_{t-1}^S + \epsilon_t \tag{8}$$

At this point they assume $\alpha_1 = 1$ and $\alpha_0 = 0$ (accurate assumptions, based on the DGP) and rearrange terms:

$$\Delta x_t^I + x_t^S = \beta_1 x_t^S + (1 + \beta_2)x_{t-1}^S + \epsilon_t \tag{9}$$

EM&W note that with $x_t^I \sim I(1)$ and $x_t^S \sim I(0)$, equation (9) is balanced. Importantly, they note that equation (9) is not just balanced but the left-hand side is $I(0)$. They point this out as a prerequisite for the original equation (7) to be what (Pickup and Kellstedt, 2021) call “ $I(0)$ balanced”. This is the requirement outlined by Banerjee *et al.* (1993, page 167–168).

“...if the order of integration of both sides is zero (which may be insured by looking for a cointegrated set of regressors and using a sufficiently differenced term as the regressand), the T-statistics can be shown to have asymptotically normal distributions... The essential point is to find some way of reparameterizing the regression such that in the re-parameterized form, the regressors, either jointly or individually, are integrated of order zero. Correspondingly, the regressand must also be $I(0)$ ”

In other words, for the purposes of estimation and inference based on the estimated standard errors, it is not only necessary that the equation is balanced, it is also necessary that there is a re-parameterization of the empirical model in which the regressand (left-hand side) is $I(0)$ (Pickup and Kellstedt, 2021). Balance is a necessary but not sufficient condition for inference. A simple example demonstrates the issue. An empirical model that regresses $y_t \sim I(1)$ on a $x_t \sim I(1)$ is balanced:

$$y_t = \beta x_t + \epsilon_t \tag{10}$$

Equation (10) is balanced because both sides are $I(1)$ but there is no reparameterization such that the regressand and regressors (right-hand side) are integrated of order zero. In other words, this empirical model is not $I(0)$ balanced. As a result, if the two variables are unrelated, the estimation of (10) results in the well-known spurious regression problem (Granger and Newbold, 1974).

EM&W claim that their empirical model is balanced and $I(0)$ balanced and therefore provides correct inference regarding the true relationship between y_t and x_t^S . The first thing to note is that the transformation used by EM&W is unnecessary to demonstrate that their empirical model (7)

⁶It also assumes an infinite long-run effect for Δx_t^S on y_t .

is balanced. It is balanced because the regressand is $I(1)$ and the regressors are in combination $I(1)$. The regressors are a linear combination of a $y_{t-1} \sim I(1)$ variable, two $x_t^S \sim I(0)$ variables, and an $I(0)$ error term.⁷ This is important because while EM&W are absolutely right when they say we “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” (p. 19), they seem to imply they have demonstrated that you can have a balanced empirical model where the right-hand side is entirely stationary and the left-hand side is an $I(1)$ variable (e.g., p. 2): “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, i.e., integrated, components. That is, 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 is not what they have demonstrated. Their empirical model is $I(1)$ on both sides. This is why it is balanced. To be clear, EM&W clearly recognize that it is the presence of $y_{t-1} \sim I(1)$ on the right-hand side that balances the equation. It is the demonstration of why it is balanced that is incorrect. It is this and the way in which they describe their results that could lead the reader to an incorrect conclusion about how to determine balance and what is permissible.

The second thing to note is that EM&W’s empirical model is not actually $I(0)$ balanced. Equation (9) is not a reparameterization of equation (7). The logic EM&W use to transform (7) into (9) means the two models are not isomorphic and the $I(0)$ balance of one cannot be used to demonstrate the $I(0)$ balance of the other. As noted previously, they substitute the DGP for Y_t into (7), and assume $\alpha_1 = 1$ and $\alpha_0 = 0$. If equation (9) was a reparameterization of equation (7), we could define some function ϕ that when applied to the variables in (7) would produce (9). However, no such function exists because setting $\alpha_1 = 1$ and $\alpha_0 = 0$ is not a reparameterization. They are restrictions. They might be valid but they would need to be placed on the empirical model before estimation. Without doing so, (7) is not $I(0)$ balanced.

If we wished to estimate a model that is $I(0)$ balanced, we could make the restrictions $\alpha_1 = 1$ and $\alpha_0 = 0$, and estimate:

$$\Delta y_t = \beta_1 x_t^S + \beta_2 x_{t-1}^S + \epsilon_t \tag{11}$$

This is the partial first differenced model and EM&W also estimate this model. It consistently estimates $\beta_1 = 1$ and $\beta_2 = -1$, which is an unbiased estimate of the relationship in the DGP (5):

$$\Delta y_t = \Delta x_t^S + \epsilon_t \tag{12}$$

But, because they would like to demonstrate that the ARDL/GECM can provide an unbiased estimate, EM&W also estimate the original empirical ARDL model (7) and the GECM version of it. They argue that these are equivalent to the partial first differenced model because $\alpha_1 = 1$ in the DGP for Y_t . However, they are only truly equivalent once you place this restriction on the empirical models. Just because you could hypothetically place the restrictions on the models does not make them equivalent. To see why the distinction is important, consider applying the logic proposed by the authors to the following example.

Let the DGP be:

$$\begin{aligned} y_t &= \alpha_0 + \beta_1 x_{t-1} + v_t \\ x_t &= x_{t-1} + \epsilon_t \end{aligned} \tag{13}$$

where $y_t \sim I(0)$ and $x_t \sim I(1)$, $\beta_1 = 0$, $\text{corr}(\epsilon_{t+j}, v_t) = 0$ for all $j \neq 0$ and $\text{corr}(\epsilon_t, v_t) \neq 0$. This is the setup for the spurious regression problem described by Mankiw and Shapiro (1985, 1986), which motivates the discussion of equation balance in Banerjee *et al.* (1993). In the spurious regression

⁷Note EM&W claim to have excluded x_t^I from the RHS but it is actually there in the lag of y_t .

problem described by Mankiw and Shapiro (1985; 1986), the further $corr(\epsilon_t, v_t)$ is from zero, the greater will be the rejection rate of the null hypothesis $\beta_1 = 0$.⁸ At the 5 percent level, the rejection rate quickly departs from 5 percent.

Clearly equation (13) is not balanced. The left-hand side is $I(0)$ and right-hand side is $I(1)$. As Banerjee *et al.* (1993) point out, this lack of balance is the source of the spurious regression problem described by Mankiw and Shapiro (1985; 1986). Using the logic applied by EM&W, we can hypothetically constrain $\beta_1 = 0$, since this is true in the DGP. As a result (13) becomes:

$$y_t = \alpha_0 + \epsilon_t \tag{14}$$

and we have achieved balance (and $I(0)$ balance). By the logic applied by EM&W, the estimation of (13) is equivalent to the estimation of (14), in which case the Mankiw and Shapiro (1985; 1986) problem is not a problem. Clearly, that is incorrect. Note that if the restriction is actually placed on the model and we estimate (14), then our equation really is $I(0)$ balanced but this is not what EM&W are proposing. EM&W’s empirical model is only $I(0)$ balanced if we make the restriction $\alpha_1 = 1$ before estimation. This is consistent with the long known fact that when estimating a first-order autoregressive model with a unit coefficient on the LDV (i.e., $\alpha_1 = 1$), one cannot assume the test statistics will have the usual limiting distribution and if one knew the coefficient was 1, one should apply the restriction before estimation (Anderson, 1959).

Due to the lack of $I(0)$ balance in the ARDL/GECM (given the DGP), the results can sometimes be spurious. As Philips’ simulations demonstrate, this is empirically only sometimes true, and as EM&W argue this stops being the case as T becomes very large (see also Enns and Wlezien, 2017). Notably, the problems seem to be greater for the estimation of the long-run effect (which includes the coefficient on the LDV $\frac{\beta_1 + \beta_2}{(1 - \alpha_1)}$) than the short-run effect.⁹ Philips (2018) demonstration that the ARDL and GECM with $y_t \sim I(1)$ and $x_t \sim I(0)$ leads to substantial false positives for the estimated long-run effects. This highlights another important difference between estimating a ARDL in which $\alpha_1 = 1$ in the DGP and estimating a model in which α_1 is restricted to 1 prior to estimation, such as in equation (6). In the first instance, it is assumed that x_t has some long-run effect on y_t and that the long-run effect is part of the estimation, leading to Type 1 errors. In the second instance, the long-run effect is correctly assumed to be 0 prior to estimation.

In the previous section, I explained why I disagreed with EM&W’s explanation for why the performance of the ARDL/GECM improves as T increases, leaving unanswered the question of why the lack of $I(0)$ balance becomes less of an issue for both the short-and long-run effects. I believe the answer may lay in the fact that while the constraint $\alpha_1 = 1$ needs to be placed on the empirical model in order to achieve $I(0)$ balance, if there is enough information in the data such that the estimate for α_1 is very close to 1, this may be sufficient¹⁰.

To demonstrate this, I use the DGP from EM&W (equation (2) above) and simulate 100,000 data sets for each value of T from 25 to 500 (in increments of 5).¹¹ I then apply the empirical model used by EM&W—the ARDL—to this data. As a point of comparison, I also apply the empirical model that best fits the DGP (equation (6)). This baseline (first difference) model regresses Δy_t on Δx_t^S . I then adjust the DGP so that there is no effect of x_t^S on y_t and repeat the exercise. I do this to show how the reduction in Type 1 and 2 errors as T increases corresponds to an increasingly precise estimate of $\alpha_1 = 1$ (Figure 1, top left panel). Note the first-difference (FD) model assumes the long-run effect of x_t^S on y_t is 0, which is the true value in both DGPs. Therefore, there are no Type 1 or 2 errors to report for the long-run effect from

⁸Banerjee *et al.* (1993) point out that this is not a problem of simultaneity bias, because the regressor x_{t-1} is uncorrelated with v_t .

⁹This is based on the use of a nonlinear Wald statistic for inference on the long-run effect.

¹⁰Note though that there is a downward bias on the estimation of α when it is equal to 1.

¹¹I set $\rho = 0.8$

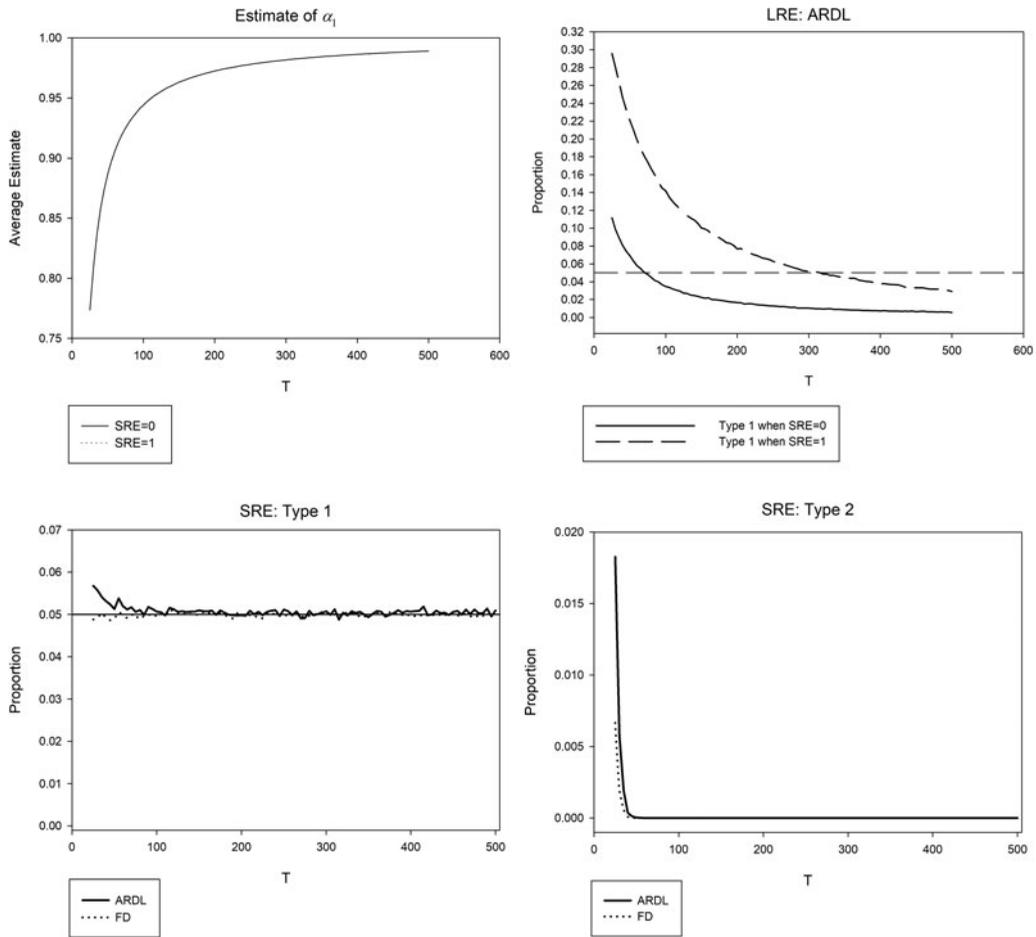


Fig. 1. Type 1 and 2 errors as $\hat{\alpha}_1 \rightarrow 1$.

the first-difference model (because it is assumed to be 0), and there are no Type 2 errors to report for the long-run effect estimated by the ARDL (because the true value is always 0). However, we can report the long-run Type 1 errors for the ARDL both when the short-run effect is 0 and when it is 1.

The results show that at low values of T , the ARDL exhibits a higher rate of Type 1 and 2 errors, compared to the baseline. The results also show that as T increases, the estimate of α_1 approaches 1 and the performance of the ARDL becomes equivalent to that of the baseline model for the short-run effect, and the Type 1 errors for the long-run effect drop to (or even below) expected levels. In other words, as the estimate of α_1 approaches the value of the constraint that would need to be placed in order to achieve $I(0)$ balance, the performance of the ARDL approaches optimality.

It is worth noting that the Type 1 and 2 errors for the short-run effect are not very large, even at small T . If the research practitioner is not interested in estimating long-run effects, the ARDL (and GECM) may produce reasonable estimates, even if it is not $I(0)$ balanced. However, the first difference model (6), which is $I(0)$ balanced (given the DGP), is the better model. Also, as the estimation of α_1 approaches 1, the estimation of the long-run effect in the ARDL/GECM will approach infinity and at $\hat{\alpha}_1 = 1$, the estimate is undefined. This is another drawback of using the ARDL/GECM when the DGP is as EM&W define it.

Overall, I agree with EM&W that the ARDL/GECEM provides reasonable inference with the DGP that they have defined (normally distributed errors, etc.), when T is large (although note that at any T , it can periodically produce very large estimates for the long-run effect when the true value is 0). I disagree with the reason proposed by EM&W and therefore disagree that the ARDL/GECEM performs well when T is not large. I also disagree that the ARDL/GECEM is $I(0)$ balanced given their DGP. This is an interesting example of a model that is not $I(0)$ balanced but may perform well under certain circumstances (estimating the short-run effect with a moderately large T). As Sims *et al.* (1990) note, there are instances in which the limiting distribution of the test statistics for some of the coefficients in a model will have standard distributions, even when the condition of $I(0)$ balance is not met.¹² This may not necessarily hold for all regressors or combinations (linear or otherwise) of these regressors. Sims (1978) gives an example where the DGP is:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t \quad (15)$$

where $\epsilon_t \sim I(0)$, $y_t \sim I(1)$, and $\alpha_1 + \alpha_2 = 1$. He demonstrates that if we estimate (15) by OLS, we are justified in testing $\alpha_1 = 0$ and/or $\alpha_2 = 0$ with the usual t -statistic. However, the F -statistic for $\alpha_1 + \alpha_2$ has a nonstandard distribution and so we are not justified in using the usual critical values. Without $I(0)$ balance it can be difficult to know when the test statistics will and will not have standard distributions. Ultimately, it is best if the research practitioner can estimate a model that is $I(0)$ balanced.

3. What next?

A couple of things are revealed by the above discussion. First, there are instances when a model is balanced but not $I(0)$ balanced that the usual test statistics are appropriate for inference. However, there is little understanding within political science of when this is the case. Further, even when the usual test statistics are appropriate for some of the parameters in the model, they may not be for other parameters or for combinations of the parameters. Given this, the best advice for the average research practitioner is to avoid estimating a model that is not $I(0)$ balanced. This leads to the second thing revealed by the above discussion.

There remains a lack of clarity on how to determine if a model is balanced and $I(0)$ balanced. There is agreement on the definition of balance but disagreement on how to determine balance and what constitutes a reparameterization to determine if a model is $I(0)$ balanced. A clear and accessible exposition on how to determine balance and how to seek a reparameterization to determine $I(0)$ balance would be a service to the discipline—especially, if it does these things for models generally, beyond just the ARDL/GECEM.

This is a more difficult task than it may at first appear. As Banerjee *et al.* (1993, page 192) note, “it is necessary to keep track of the orders of integration of both sides of the regression equation.” This means it is important to think about the order of integration of each variable, and their combination. This is something that political scientists often ignore and even when they do not, there is lack of clarity on how to do this. For example, the debate over the use of the GECEM has made much of its use in Volsho and Kelly (2012). At times, the *Democratic president* variable used in the model has been argued to be stationary. The variable is a control for a regime shift, from a Democratic president to a Republican or vice versa. From a theoretical perspective, it is not clear that a deterministic variable like this can be considered stationary or integrated. There is also little discussion within political science of variables having a higher order of integration than 1, or of the concept of multi-cointegration. There are also variables that have fractional orders of integration and non-linear variables that have no order of integration (Berenguer-Rico and Gonzal,

¹²In an earlier version of their contribution to the symposium, KK&L enumerated a number of these examples.

2013). The application of the concepts of balance and $I(0)$ balance requires an understanding of how to account for such variables.

Finally, all published discussions of balance to date have been in the context of single-equation time series. As a concept, balance is equally applicable to multi-equation time series and panel data. Applying the lessons learned from single-equation time series to multi-equation time series might be straightforward but it might not be. Certainly, the application to panel data will come with complications both conceptually and (maybe even more so) in practice. Work in this area would constitute a major contribution to the field.

Data. For Dataverse replication materials, see <https://doi.org/10.7910/DVN/IITPH8> (Pickup, 2022).

Acknowledgments. I thank the authors of the articles in this symposium, the reviewers, and the Journal Editor. All errors are my own.

References

- Achen CH (2005) Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science* **22**, 327–339.
- Anderson TW (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. *The Annals of Mathematical Statistics* **30**, 676–687.
- Banerjee A, Dolado J, Galbraith JW and Hendry D (1993) *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford, UK: Oxford University Press.
- Berenguer-Rico V and Gonzal J (2013) Summability of stochastic processes—a generalization of integration for non-linear processes. *Journal of Econometrics* **178**, 331–341.
- Cochrane J (1991) A critique of the application of unit root tests. *Journal of Economic Dynamics and Control* **15**, 275–284.
- Enns PK and Wlezien C (2017) Understanding equation balance in time series regression. <https://thepoliticalmethodologist.com/2017/06/23/understanding-equation-balance-in-time-series-regression/>.
- Freeman JR (2016) Progress in the study of nonstationary political time series: a comment. *Political Analysis* **24**, 50–58.
- Granger CW and Newbold P (1974) Spurious regressions in econometrics. *Journal of Econometrics* **2**, 111–120.
- Johansen S (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**, 231–254.
- Keele L, Linn S and Webb CM (2016a) Concluding comments. *Political Analysis* **24**, 83–86.
- Keele L, Linn S and Webb CM (2016b) Treating time with all due seriousness. *Political Analysis* **24**, 31–41.
- Lebo MJ and Grant T (2016) Equation balance and dynamic political modeling. *Political Analysis* **24**, 69–82.
- Mankiw N and Shapiro M (1985) Trends, random walks and tests of the permanent income hypothesis. *Journal of Monetary Economics* **16**, 165–74.
- Mankiw N and Shapiro M (1986) Do we reject too often? Small sample properties of tests of rational expectations models. *Journal of Monetary Economics* **120**, 139–45.
- Pesaran M, Shin Y and Smith R (2001) Balance testing approaches to the analysis of level relationships. *Journal of Applied Econometrics* **16**, 289–326.
- Philips AQ (2018) Have your cake and eat it too? Cointegration and dynamic inference from autoregressive distributed lag models. *American Journal of Political Science* **62**, 230–244.
- Pickup M (2022) Replication for: equation balance in time series analysis: lessons learned and lessons needed. <https://doi.org/10.7910/DVN/IITPH8>.
- Pickup M and Kellstedt P (2021) Balance as a pre-estimation test for time series analysis. *Political Analysis* 1–10. <https://doi.org/10.1017/pan.2022.4>.
- Sims C (1978) Least-squares estimation of autoregressions with some unit roots, discussion paper.
- Sims C, Stock J and Watson M (1990) Inference in linear time series models with some unit roots. *Econometrica* **58**, 113–144.
- Stock JH and Watson MW (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* **61**, 783–820.
- Volsho T and Kelly N (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* **5**, 679–699.
- Webb C, Linn S and Lebo M (2019) A bounds approach to inference using the long run multiplier. *Political Analysis* **27**, 281–301.
- Wooldridge J (2010) *Econometric Analysis of Cross Section and Panel Data*. 2nd ed., Cambridge, MA: MIT Press.