Political Analysis (2016) 24:83–86
doi:10.1093/pan/mpv030

Concluding Comments

Luke Keele and Suzanna Linn
Department of Political Science, Pennsylvania State University, State College, PA 16802
e-mail: ljk20.psu.edu (corresponding author) and slinn@la.psu.edu

Clayton McLaughlin Webb
Department of Political Science, University of Kansas, Lawrence, KS 66049
e-mail: webb767@ku.edu

Edited by Janet Box-Steffensmeier

This issue began as an exchange between Grant and Lebo (2016) and ourselves (Keele, Linn, and Webb 2016) about the utility of the general error correction model (GECM) in political science. The exchange evolved into a debate about Grant and Lebo's proposed alternative to the GECM and the utility of fractional integration methods (FIM). Esarey (2016) and Helgason (2016) weigh in on this part of the debate. Freeman (2016) offers his views on the exchange as well. In the end, the issue leaves readers with a lot to consider. In his comment, Freeman (2016) argues that the exchange has produced little significant progress because of the contributors' failures to consider a wide array of topics not directly related to the GECM or FIM. We are less pessimistic. In what follows, we distill what we believe are the most important elements of the exchange—the importance of balance, the costs and benefits of FIM, and the vagaries of pre-testing.

1 Balance

We believe the discussion of equation balance was an important part of the initial exchange. Balanced regressions are those in which the regressand is of the same order of integration as the regressors, or any linear combination of the regressors (Bannerjee et al. 1993). For a time-series equation to be balanced, the regressand and the regressors must all be stationary or a combination of the nonstationary regressand and one or more nonstationary regressors must constitute a cointegrating vector.

Freeman argues that we have overstated the importance of equation balance. He points out that it is possible to derive some inferences from unbalanced models if one uses the correct distributions for the test statistics. Freeman's point about test statistics is correct. Appropriate limiting distributions can be derived for hypothesis tests in unbalanced equations. Bannerjee et al. (1993) note that there are special circumstances where certain inferences can be drawn from unbalanced regressions. One example Bannerjee et al. (1993, 166) give is the Dickey Fuller test. While Freeman's point is technically correct, we believe it is important for readers to consider this point in a broader context.

There is no evidence that analysts can derive reliable inferences about long-run relationships from unbalanced models. Bannerjee et al. (1993, 167) say that attempts to explain $I(0)$ variables with variables integrated of higher orders fail because these variables diverge by ever-larger amounts over time. Enders (2015, 199) says these kinds of regression equations are "meaningless." Outside of a few special circumstances, the degree to which reliable inferences about short- and long-run effects can be derived from unbalanced models is not well understood. How unbalanced is too unbalanced? What types of violations are acceptable? We believe analysts should be wary of specifying unbalanced models until these questions are answered.

2 Fractional Integration Methods

The second important feature of the exchange relates to how often analysts should use fractional integration methods (FIMs). Grant and Lebo (2016) argue that most political science data are fractionally integrated and that FIMs are necessary for safe inferences. We cautioned that considerable uncertainty surrounds estimates of the fractional differencing parameter, d. We acknowledged that FIMs are optimal when we know the true data-generating process, but cautioned that these methods are less than optimal in a variety of circumstances.

The debate about FIMs raises two questions. The first question is the subject of both Esarey (2016) and Helgason (2016): if we ignore that data are fractionally integrated and (potentially) cointegrated, what are the costs of using traditional time-series regression models like the autodistributed lag (ADL) or general error correction model (GECM)? Esarey (2016) examined the quality of inferences from autoregressive distributed lag models when the time series of interest are (stationary) fractionally integrated processes. He considers cases where the data are cointegrated and cases when they are independent processes.[1] Esarey concludes that in many cases "an ADL/ECM provides a serviceable approximation in a short T data set, where d is inaccurately estimated and overfitting is a concern." We think this is added evidence of the flexibility of these methods.

Helgason (2016) compares FIMs to the GECM when the data are fractionally integrated. He considers both stationary and nonstationary values of d. He finds results similar to Esarey. The GECM underestimates long-run effects (but not short-run effects) when data are fractionally integrated. He also finds that FIMs often do a better job approximating long-run effects. However, we think there are important caveats to Helgason's findings. When sample sizes are small and short-run dynamics are present in the individual time series, "the GECM provides more accurate predictions than FIMs and is less prone to produce wildly inaccurate predictions."[2]

According to Esarey and Helgason, the chief *potential* cost when adopting traditional time-series methods for fractionally integrated series is an underestimation of the long-run effect of X_t on Y_t. Even when traditional methods do underestimate the long-run effects, the differences are minor. Whether this potential cost is acceptable depends on the goals of the analysis, the size of the sample, and whether short-run dynamics exist. Conservative estimates are generally preferable to overestimated effects, sample sizes are considerably smaller than 1500 in much applied work, and it remains unclear how commonplace short-run dynamics are in conjunction with fractional integration. Thus, we agree with Esarey that the ADL/GECM offers a "serviceable approximation" for analysts, even when data are fractionally integrated.

The second question raised by this debate, and still unanswered, is how do we go wrong if we use FIMs to model stationary, $I(0)$, or integrated, $I(1)$, time series? It seems likely that there is a potential to overestimate long-run effects if we incorrectly assume $I(0)$ data are fractionally integrated, given the longer memory implied by fractional integration, and underestimate it if the data are $I(1)$. It also seems likely that errors will be larger in smaller samples and for stationary processes with higher-order dynamics. At present, to our knowledge, these questions remain unanswered in the time-series literature. It may be that the costs are high enough relative to the costs of incorrectly assuming data are $I(0)$ or $I(1)$ that FIMs are not a tractable alternative until sample sizes are quite large. However, it may be that the costs of using FIMs are minimal. To fully assess the relative efficacy of the GECM (or ADL) and FIM, we need an answer to this second question as well. This serves as an additional avenue for future research.

3 Back to Pretesting

Freeman (2016) raises several points that go beyond the issues at the center of this exchange. In particular, Freeman would have us develop a formal pretest design, assess the prevalence of different types of univariate dynamics characterizing political time series, elaborate tests for weak

[1]In both Esarey (2016) and Helgason (2016), as well as Grant and Lebo (2016) and our own work, consideration is limited to independent variables that are weakly exogenous for the parameters of interest. If X_t is not weakly exogenous, then other methodologies, such as vector autoregression, must be employed.

[2]And of course our uncertainty over the size of short- and long-run effects using FIMs is understated given the inherent need to estimate the model in two stages: the uncertainty over the estimate of d, which we showed is understated, and the uncertainty inherent in the GECM.

exogeneity, and discuss the implications of our work for theory building. Moving somewhat further afield, he would also like to see a consideration of the meaning and implications of balance for panel design as well as how Bayesian time-series analysis could help us wrestle with questions relating to the proper characterization of univariate time-series dynamics. These are laudable goals and represent an agenda for future research. Here, we briefly consider two of these points: development of a formal pretest design and characterization of the "true" nature of univariate dynamics of political time series.

Time-series texts offer a number of tests for inferring the dynamics of univariate time series. The general advice is that analysts should estimate a number of tests and proceed based on the preponderance of the evidence. Unfortunately, there will invariably be cases where this strategy will not definitively settle the question of the "true" dynamic nature of a specific time series. We can weigh the evidence and more or less confidently infer dynamics, but the task is complicated by at least three issues. First, any pretest design for determining the true DGP is based on tests with low power against local alternatives. For example, we often fail to reject both the null hypothesis that time series are stationary and that they are unit roots. Second, the patterns of autocorrelation in the series often approximate multiple characterizations of the DGP. Any given ARFIMA(p,d,q) process, for example, has an ACF arbitrarily similar to an ARMA(p,q) process within sample. Third, potential structural breaks, particularly of unknown timing, can confound pretesting. The mixed signals that often result from pretesting applied to a single series, or the same series over different time periods, speak volumes. Even a careful, conscientious pretesting strategy may be unable to arbitrate clearly among the alternatives in a number of cases. Analysts should apply several tests and use their best judgment.

4 Conclusion

In closing, we believe there are several useful conclusions that can be drawn from this exchange. First, balance matters. The cases identified by Grant and Lebo (2016) demonstrate that estimating unbalanced equations can lead to incorrect inferences. However, we maintain that the GECM remains a sound estimation strategy when equations are balanced. Second, while FIMs are sound in theory, the application of these methods may be problematic in practice. Third, based on our evidence, and that presented by Esarey (2016) and Helgason (2016), traditional time-series methods represent reasonable alternatives to FIMs in many cases. Taken together, the exchange suggests analysts should be cautious about throwing workhorse models like the ADL and the GECM to the wind.

The exchange suggests several best practices for applied time-series analysis. Analysis should begin with a well-developed theory. Analysts should select an identification strategy, collect data, and pretest to infer the dynamic processes underlying the individual time series. Analysts should be aware that such tests have low power to distinguish between local alternatives and that failing to consider structural breaks can lead to false inferences. Next, analysts should choose a general model consistent with the theory and the properties of the data, estimate the model, and test restrictions. Variables may need to be transformed or filtered, as suggested by table 5 in Keele, Linn, and Webb (2016). The final model should be balanced, and there should be no residual autocorrelation. Analysts should take care not to overfit the data. Good empirical work is hard to mass produce, and the modeling of dynamic relationships is no exception.

Conflict of interest statement. None declared.

References

Bannerjee, Anindya, Juan Dolado, John W. Galbraith, and David F. Hendry. 1993. *Integration, error correction, and the econometric analysis of non-stationary data*. Oxford: Oxford University Press.

Enders, Walter. 2015. *Applied econometric time series*. 4th ed. New York: Wiley and Sons.

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

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

Grant, Tayler, and Matt Lebo. 2016. Error correction methods with political time series. *Political Analysis* 24:3–30.
Helgason, Agnar Freyr. 2016. Fractional integration methods and short time series: Evidence from a simulation study. *Political Analysis* 24:59–68.
Keele, Luke J., Suzanna Linn, and Clayton McLaughlin Webb. 2016. Treating time with all due seriousness. *Political Analysis* 24:31–41.