Political Analysis (2016) 24:59–68
doi:10.1093/pan/mpv026

Fractional Integration Methods and Short Time Series: Evidence from a Simulation Study

Agnar Freyr Helgason

Social Research Centre, University of Iceland, Gimli G-201, Sæmundargötu 2, 101 Reykjavík, Iceland
e-mail: afh@hi.is (corresponding author)

Edited by Janet Box-Steffensmeier

Grant and Lebo (2016) and Keele, Linn, and Webb (2016) provide diverging recommendations to analysts working with short time series that are potentially fractionally integrated. While Grant and Lebo are quite positive about the prospects of fractionally differencing such data, Keele, Linn, and Webb argue that estimates of fractional integration will be highly uncertain in short time series. In this study, I simulate fractionally integrated data and compare estimates from the general error correction model (GECM), which disregards fractional integration, to models using fractional integration methods over thirty-two simulation conditions. I find that estimates of short-run effects are similar across the two models, but that models using fractionally differenced data produce superior predictions of long-run effects for all sample sizes when there are no short-run dynamics included. When short-run dynamics are included, the GECM outperforms the alternative model, but only in time series that consist of under 250 observations.

1 Introduction

Political science has seen a boom in the use of the general error correction model (GECM) for time-series analysis in recent years. While this has been a welcome development, the GECM is not appropriate for any and all temporal data. In that vein, Grant and Lebo's (2016) contribution in this volume provides an important service by highlighting areas where practitioners have mistakenly applied and misinterpreted results based on the GECM. Similarly, Keele, Linn, and Webb's (2016) contribution provides a much-needed discussion of the interpretation of the results of the GECM, as well as suggestions for how to proceed based on the univariate properties of time series under study.

While there is some consensus between the two studies, they provide diverging recommendations in other areas. Where the dialogue is perhaps least helpful is in providing guidelines for practitioners working with short time series. This is unfortunate, since short time series are quite common in political science and are often analyzed with the GECM. Given that the present dialogue arose in response to the mistaken application of the GECM, it is doubly important to provide clear recommendations for practitioners working with "typical" data.

One of the primary issues of contention concerns the applicability of fractional integration methods to short time series. Grant and Lebo are quite positive about the prospects of fractionally differencing such data, suggesting that estimates of d are unbiased "even with N as small as 40" (Grant and Lebo 2016, 26), with the implication being that fractional differencing can successfully be applied to relatively short time series. Meanwhile, Keele, Linn, and Webb offer a more reserved assessment. While they do agree that testing for fractional integration and fractionally differencing or using a fractional error correction model is appropriate for relatively long time series, they offer a pragmatic limitation for the use of the method with shorter time series: Estimating fractional

Authors' note: The author thanks Janet Box-Steffensmeier for comments. Replication materials are available online as Helgason (2016).

integration in short time series is a highly uncertain endeavor, with the implication being that the method is not a feasible alternative for such data.

Keele, Linn, and Webb provide simulation evidence to support their claim, that is, that in short time series the estimate of fractional integration is highly uncertain. However, in their simulation analysis they do not provide a comparison with the most relevant counterfactual: namely, whether assuming that data are stationary or integrated while they are actually fractionally integrated is preferable to employing an uncertain estimate of fractional integration and proceeding accordingly.

This is a critical point, both for empirical and theoretical reasons. Empirically, there is considerable evidence that many commonly used time series in political science are fractionally integrated (e.g., Box-Steffensmeier and Smith 1996; Byers, Davidson, and Peel 2000; Lebo, Walker, and Clarke 2000), while theoretically, we often have good reason to assume that time series in political science are fractionally integrated (Granger 1980; Box-Steffensmeier and Tomlinson 2000; Lebo and Weber 2015). The applicability of fractional integration methods to short time series is, thus, of considerable importance to applied researchers.

In this article, I provide some evidence as to the suitability of fractional integration methods for common sample sizes in political science. I generate two fractionally integrated variables x and y and consider four potential bivariate relationships between them: that there is no relationship between the two variables, and either $d_x = d_y = 0.4$ or $d_x = d_y = 0.8$, or that there is a long-run cointegrating relationship between x and y, and either $d_x = d_y = 0.4$ or $d_x = d_y = 0.8$.[1] I simulate data for samples of sizes 50, 100, 250, or 1000, and consider the possibility that each of the variables follow an ARFIMA(0,d,0) or ARFIMA(1,d,0) process. Finally, I compare estimates of the short-run and long-run effects of x on y from a GECM which doesn't consider the possibility that the data are fractionally integrated, and an ECM after fractional differencing, which furthermore considers the potential for fractional cointegration. I refer to the former strategy as GECM, while I refer to the latter strategy as FIM for the remainder of the article.

The results indicate that the choice of estimation strategy depends on the quantity of interest and whether or not there are short-run dynamics at play. If the analyst is primarily interested in short-run effects, the two models perform similarly across most simulation conditions. However, if the analyst is primarily interested in long-run effects, the relative performance of the two methods differs based on whether short-run dynamics are present. If no short-run dynamics are included, the FIM performs equally well or outperforms the GECM, with the benefits of using FIM increasing with sample size. However, if there are (unknown) short-run dynamics, the relative performance of the models differs by sample size. With short time series ($T < 250$), short-run dynamics interfere with the estimation of fractional integration and the GECM provides marginally better predictions of the long-run effect of x on y. However, as the sample size grows, testing for fractional integration and fractionally differencing data becomes the better strategy to follow.

2 Simulation Design

To compare the performance of the GECM and the FIM with fractionally integrated data, I simulate data for two variables, y and x.[2] Following Nielsen (2004, 2010), the data-generating process is based on the triangular model

$$y_t - \beta x_t = (1 - L)^{-(d-b)} e_{1t,} \tag{1}$$

$$x_t = (1 - L)^{-d} e_{2t,} \tag{2}$$

where $t = 1, 2, \ldots T$ indexes time period, L is the lag operator, β is the relationship between x and y, and (e_{1t}, e_{2t}) are $I(0)$ processes, possibly with short-run dynamics. Throughout the study I assume

[1] I limit the analysis to cases in which x and y are fractionally integrated of the same order, since these are cases in which analysts are likely to erroneously conclude that the two variables are either both stationary or both contain a unit root.
[2] Replication materials are available online as Helgason (2016).

that e_{1t} and e_{2t} are uncorrelated, so that x is weakly exogenous to y, by design. The fractional integration order of x is d (denoted $I(d)$), while the fractional integration order of the linear combination $y_t - \beta x_t$ is d–b (denoted $I(d-b)$). If $\beta \neq 0$ and $d > b > 0$, the two variables are cointegrated. I consider several alternative parameter combinations for the above model, varying β, d, b, T, and the properties of (e_{1t}, e_{2t}).

2.1 *Cointegrating Relationship*

I consider four combinations of β, d, and b, which determine the relationship between x and y. In each of the cases, I set $d=0.4$ or $d=0.8$, which matches two of the five specifications considered by Keele, Linn, and Webb (2016, 18). I chose the value of b so that it covers cases in which there is no cointegrating relationship between the variables ($b=0$), where there is a cointegrating relationship between them and a linear combination of the two variables is stationary ($b=d$), and in which a cointegrating relationship between them exists, but their linear combination is itself fractionally integrated ($d > b > 0$).

1. $\beta=0$, $d=0.4$, and $b=0$. In this case, there is no cointegration relationship between the two variables. Thus, x and y are both $I(0.4)$ variables and there is no linear combination between them that is integrated of a lower order than d. Since $d<0.5$, the variables are mean reverting with finite variance.
2. $\beta=0$, $d=0.8$, and $b=0$. This case is identical to the above case, except $d>0.5$, so the variance of the series is not finite.
3. $\beta=0.5$, $d=0.4$, and $b=0.4$. In this case, there is a cointegrated relationship between the two variables. Thus, while x and y are both $I(0.4)$, the linear combination $y-0.5x$ is $I(0)$.
4. $\beta=0.5$, $d=0.8$, and $b=0.6$. This case is identical to the above case, except that the variance of the two series is not finite ($d>0.5$) and the linear combination of x and y is itself fractionally integrated of order $I(0.2)$.

2.2 *Sample Size*

I consider four possible sample sizes, $T = 50, 100, 250, 1000$. Sample sizes of 50 and 100 are commonly found in series used in political science, a sample size of 250 is the threshold under which Keele, Linn, and Webb (2016, 27) recommend practitioners consider fitting simpler models to their data, while a sample size of 1000 or longer is more and more common in the age of data analytics.

2.3 *Short-Run Dynamics*

One of the points of disagreement between the two papers concerns the extent to which short-run dynamics complicate the estimation of fractional integration. Keele, Linn, and Webb (2016, 21) provide simulation evidence that demonstrates how shifting from an ARFIMA(0,d,0) model to an ARFIMA (p,d,q) model biases estimates of d and results in inflated rejection rates of the null hypotheses that $d=0$. Meanwhile, Grant and Lebo (2016, 25) argue that most time series used in political science are well described by a simple (0,d,0) model and that concerns over the effects of short-run dynamics are, thus, uncalled for.

Given that the two papers disagree on whether short-run dynamics are a cause for concern, I err on the side of caution and consider both processes with and without short-run dynamics below. In the baseline specification, I assume that there are no short-run dynamics present, and that the two error processes are Gaussian white noise. Formally,

$$\begin{aligned} e_{jt} &= \epsilon_{jt}, \\ \epsilon_{jt} &\sim \mathrm{NID}(0,1) \text{ for } j = 1, 2. \end{aligned} \tag{3}$$

In the alternative specification, I assume that the errors each follow an AR(1) process with $\phi = 0.5$. Formally,

$$\begin{aligned} e_{jt} &= 0.5 e_{j,t-1} + \epsilon_{jt}, \\ \epsilon_{jt} &\sim \text{NID}(0, 1) \text{ for } j = 1, 2. \end{aligned} \tag{4}$$

Altogether there are, thus, $4 \times 4 \times 2 = 32$ simulation conditions over which the GECM and FIM will be compared. I ran 500 simulations for each of the conditions, generating the series using the `ARFIMA` package in `R`.

2.4 *Estimation*

In estimating the models, I follow a simplified version of the pretesting strategy suggested in Section 5 of Keele, Linn, and Webb (2016). Below, I delineate the processes for the GECM and FIM and highlight how they differ.[3]

For the GECM, I treat series as either $I(0)$ or $I(1)$ and, thus, disregard the potential for fractional integration. If both x and y are $I(1)$ and are jointly $I(0)$, I estimate a GECM. If they are both $I(1)$ and jointly $I(1)$, I estimate a GECM in first differences. If x and y are both $I(0)$, I estimate a GECM. If one of the variables is $I(1)$ and the other is $I(0)$, I first difference the integrated variable and estimate a GECM on the transformed data. For each of the cases, I specify the model as a GECM(1,1,1), that is,

$$\Delta y_t = \alpha_1 y_{t-1} + \beta_0 \Delta x_t + \beta_1 x_{t-1}, \tag{5}$$

thus only including the immediate change in (a potentially transformed) x and the first lag of y and x. Although I concur with De Boef and Keele's (2008) recommendation that analysts should follow a general to specific modeling strategy, the GECM(1,1,1) is the model most often estimated by practitioners, and as such, provides a critical comparison.

For the FIM, I begin by fitting an ARFIMA(p,d,q) model to x, y, and the residual, $\widehat{\text{ECM}}$, from regressing y on x, using Sowell's (1992) exact maximum-likelihood estimate of d, which is implemented in the `ARFIMA` package in `R`. In keeping with the discussion on short-run dynamics above, the models considered for selection differ across treatments. When there are no short-run dynamics ($\phi = 0$), I assume the analyst proceeds according to Grant and Lebo's (2016) suggestion and disregards the potential that short-run dynamics are present. Thus, $p = q = 0$ and only the ARFIMA(0,d,0) model is considered. However, when short-run dynamics are included ($\phi = 0.5$), I assume the analyst does not know the order of short-run dynamics and estimates four models: ARFIMA(0,d,0), ARFIMA(1,d,0), ARFIMA(1,d,1), and ARFIMA(0,d,1), selecting the best model based on the Akaike information criterion (AIC). Cases with short-run dynamics, thus, also include uncertainty about whether the series have autoregressive or moving average components. I save the prewhitened residuals from the best ARFIMA model for each variable and use in all subsequent models.

I proceed based on the estimates of d for x, y, and $\widehat{\text{ECM}}$, and their respective 95% confidence intervals. If the confidence intervals of $\hat{d}_x$ and $\hat{d}_y$ overlap, I compare the intervals to the confidence interval for $\hat{d}_{\widehat{\text{ECM}}}$. If the interval for $\hat{d}_{\widehat{\text{ECM}}}$ is strictly lower than the intervals for $\hat{d}_x$ and $\hat{d}_y$, I conclude that x and y have a cointegrating relationship. Accordingly, I estimate a fractional error correction model (FECM) based on the two-step Engle–Granger strategy using the prewhitened variables, including the lag of the estimated fractional error correction term, $\widehat{\text{ECM}}_{t-1}$. Formally, I fit the model

$$\Delta^{d_y} y_t = \alpha_1 \Delta^{d_{\text{ECM}}} \widehat{\text{ECM}}_{t-1} + \beta_0 \Delta^{d_x} x_t + \beta_1 \Delta^{d_x} x_{t-1} \tag{6}$$

[3]For tractability, I do not test for structural breaks or stochastic volatility, and I assume it is known how many times a series has to be differenced to become stationary. I also do not estimate an intercept in any of the models. While these are important components of any time-series analysis, I do not expect these tests to affect the comparison between the GECM and FIM.

https://doi.org/10.1093/pan/mpv026 Published online by Cambridge University Press

to the data. If the interval for $\hat{d}_{\widehat{\text{ECM}}}$ overlaps either of the other two intervals or if the intervals for $\hat{d}_x$ and $\hat{d}_y$ do not overlap, I conclude that the variables are not cointegrated and estimate an error correction model using the prewhitened variables, that is,

$$\Delta^{d_y} y_t = \alpha_1 \Delta^{d_y} y_{t-1} + \beta_0 \Delta^{d_x} x_t + \beta_1 \Delta^{d_x} x_{t-1}. \tag{7}$$

Finally, the estimated coefficients from the GECM and FIM, as well as the selected ARFIMA models for x, y, and $\widehat{\text{ECM}}$, are saved for each of the simulated samples.

2.5 *Evaluation*

When evaluating the performance of the models above, we are primarily interested in two factors—the short-run, or immediate, effect of x on y and the long-run, or cumulative, effect of x and y. Both factors are likely to be important in political science for substantive understanding and theory testing. Capturing and comparing the short-run effect is straightforward, as the estimate of β_0 in models 5, 6, and 7 directly reports the quantity of interest. For each of the thirty-two conditions, I compare the absolute bias in the estimated coefficient

$$\text{Absolute bias} = |\hat{\beta}_0 - \beta_0|, \tag{8}$$

and the coverage of the 95% confidence interval around the point estimate (Carsey and Harden 2013).

Comparing the long-run, or cumulative, effect is not as straightforward. The coefficients of a fractionally differenced series are calculated using binomial expansion, which is infinite for fractional exponents. This means that fractionally integrated series do not have a finite long-run multiplier which can be compared across different models. Thus, when evaluating long-run effects with such data, it is necessary to truncate the forecasting period over which changes can occur and limit the estimate of the long-run effect to a finite time period (Cowpertwait and Metcalfe 2009).

Below, I report the prediction error for each of the conditions at $t = 10$, based on the model estimates.[4] To calculate the true long-run effect, I simulate a one-unit shock to x at $t = 1$ and calculate the path of y from $t = 1$ to $t = 10$ using the true parameter values and the `arfima.sim` function in `R`. For the FIM I proceed analogously, except I use estimated parameter values to simulate the path of y. For the GECM, I set $x_{t-1} = y_{t-1} = 0$ so the system is in equilibrium at $t = 0$, simulate a one-unit shock to x, and trace out the path of Δy from $t = 1$ to $t = 10$.

I report different quantities for series based on the value of d. For conditions in which $d_x = d_y = 0.8$, I report the prediction error for the value of y ten periods into the future, y_{t+10}. Formally,

$$\text{Prediction Error} = \hat{y}_{t+10} - y_{t+10}. \tag{9}$$

For conditions in which $d_x = d_y = 0.4$, I report the prediction error for the cumulative change in y ten periods into the future, $\sum_{t=1}^{10} \Delta y_t$. Formally,

$$\text{Prediction Error} = \sum_{t=1}^{10} \widehat{\Delta y_t} - \sum_{t=1}^{10} \Delta y_t. \tag{10}$$

3 Results

Figure 1 shows the results for the immediate effect of x on y for the GECM and FIM in each of the thirty-two simulation conditions. Overall, the results are remarkably consistent, with the mean absolute bias being comparable for the two models in each of the thirty-two conditions. Thus, regardless of sample size, the true relationship between the two variables, and the inclusion of

[4]While the choice of $t = 10$ is arbitrary, the results are substantively similar for other forecasting horizons. Since fractional integration primarily affects the persistence of shocks over time, there is less (more) divergence between estimates from the two models for shorter (longer) time horizons.

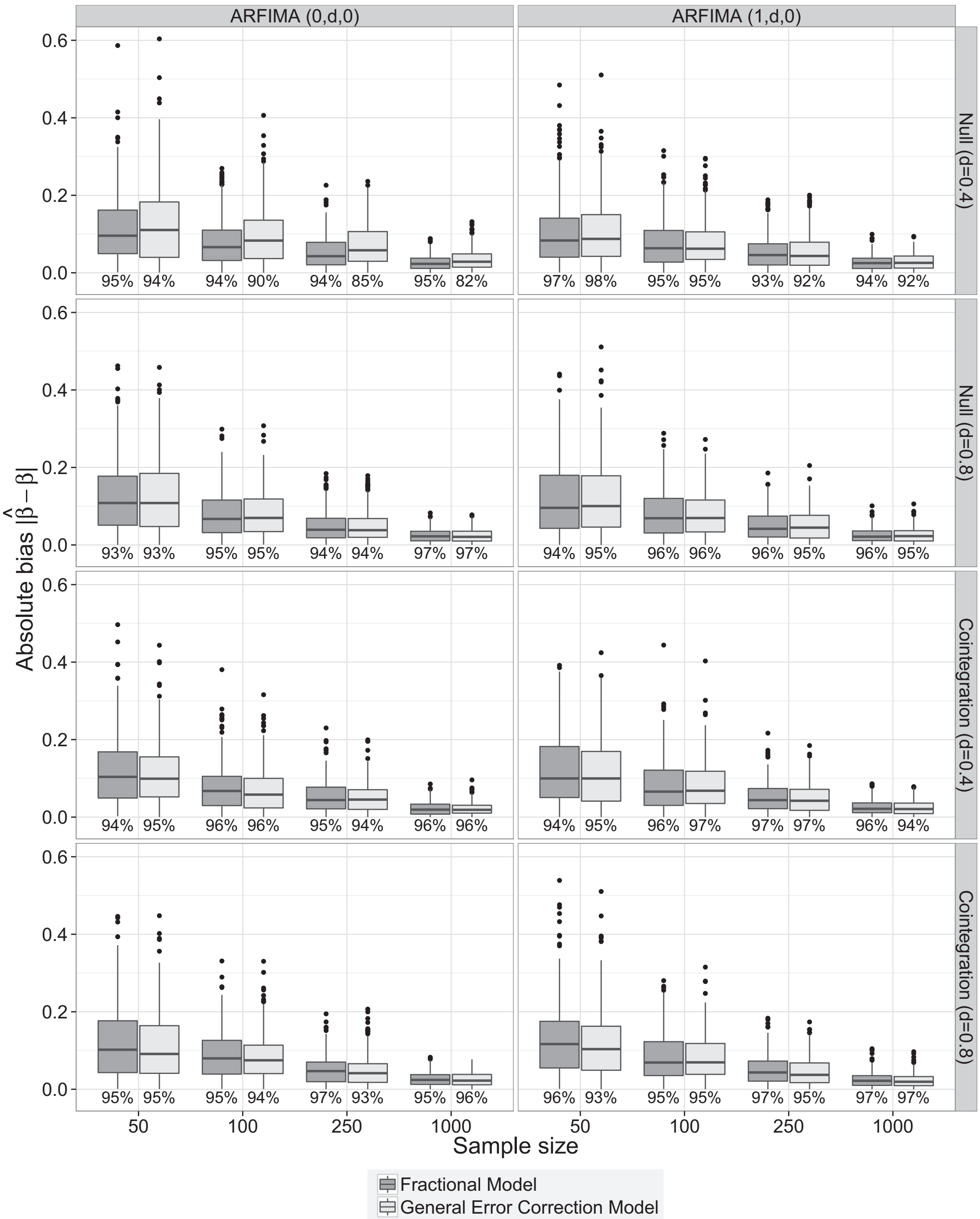

Fig. 1 Immediate effect of x on y.

Notes. Results for immediate effects of x on y for the 32 simulation conditions. Each row corresponds to one of the four cointegration relationships, while the columns correspond to whether short-run dynamics are included (right) or not (left). Within each panel, the distribution of estimates are shown for each of the four sample sizes for the GECM and FIM. The boxplots show the distribution of absolute bias—an absolute bias closer to 0 is better. Below each boxplot, the coverage probability of a 95% confidence interval is shown. A value lower than 95% indicates that the estimated standard errors are too small, while a value over 95% indicates that they are too large. Estimated coverage probabilities closer to 95% are preferable. Each model summary is based on 500 simulations.

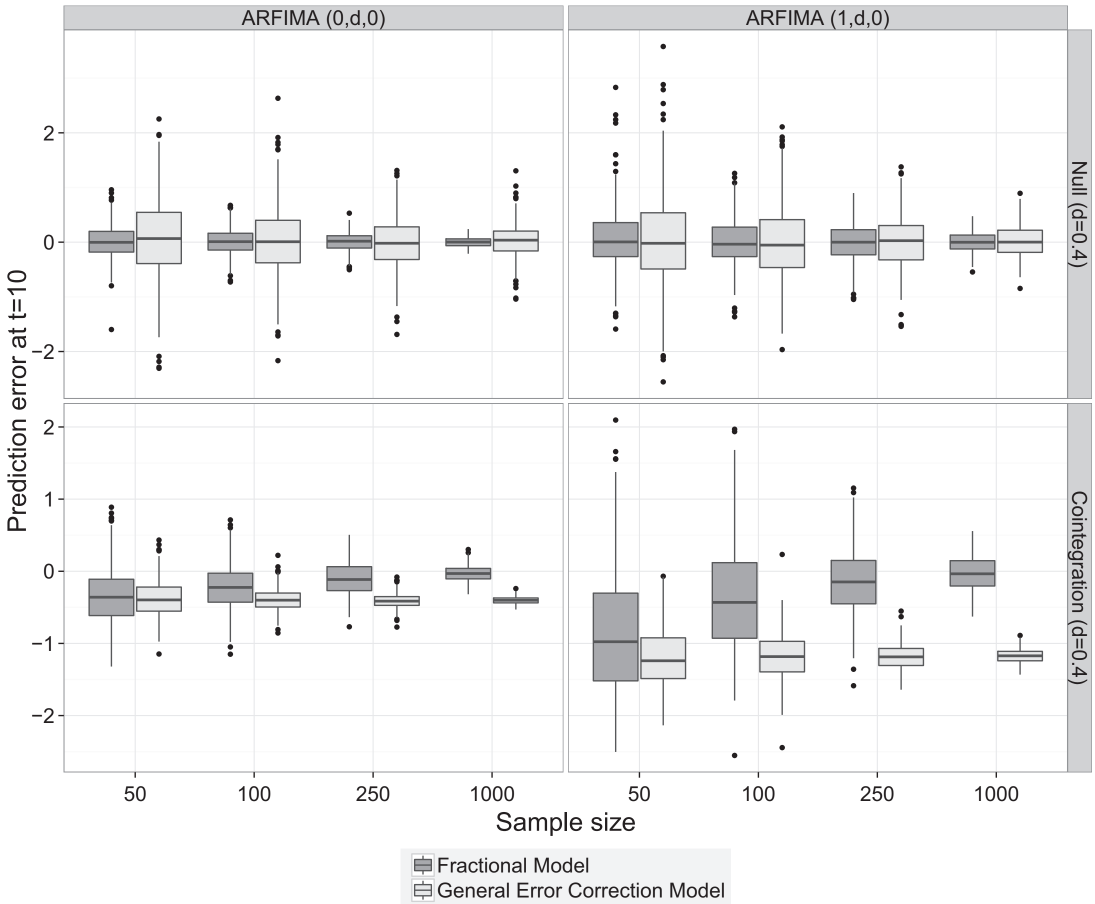

Fig. 2 Cumulative effect of x on y at $t = 10$ when $d_x = d_y = 0.4$.
Notes. Results for cumulative effect of x on y at $t = 10$ for the 16 simulation conditions in which $d_x = d_y = 0.4$ (Equation 10). The setup is based on the scheme described in Fig. 1.

short-run dynamics, the two models provide consistent and comparable estimates in terms of absolute bias. As is to be expected, the bias becomes smaller as the sample size increases, while the presence of short-run dynamics does not seem to cause more biased estimates.

A similar story can be told in terms of the coverage probabilities. For most of the conditions, the coverage probability is close to 95%, indicating that in the long run the true effect of x on y is captured in 95 out of 100 confidence intervals, which corresponds to the definition of a 95% confidence interval (Carsey and Harden 2013, 93). The only model and condition for which the coverage probabilities deviate significantly from the 95% threshold is the GECM estimates of the null relationship with $d_x = d_y = 0.4$ and no short-run dynamics (top-left panel). As the sample size increases, the coverage probability of the estimates becomes smaller, indicating that the confidence intervals estimated by the model are overly confident and that the null hypothesis of no effect is rejected too often. Since the absolute bias for the same model and condition is comparatively small, this suggests that the fault lies primarily with standard errors that are too small.

When we turn our attention to the prediction error for the long-term effect of x on y, a different story emerges. Figure 2 shows the prediction error for the cumulative effect of x on y at $t = 10$, when $d_x = d_y = 0.4$. Aside from the evaluation metric, the setup of the figure follows the same scheme as Fig. 1. Several results are of interest.

When there is no relationship between the variables and there are no short-run dynamics (upper-left panel), the FIM is considerably closer to the true cumulative effect, $\sum_{t=1}^{10} \Delta y_t = 0$, although

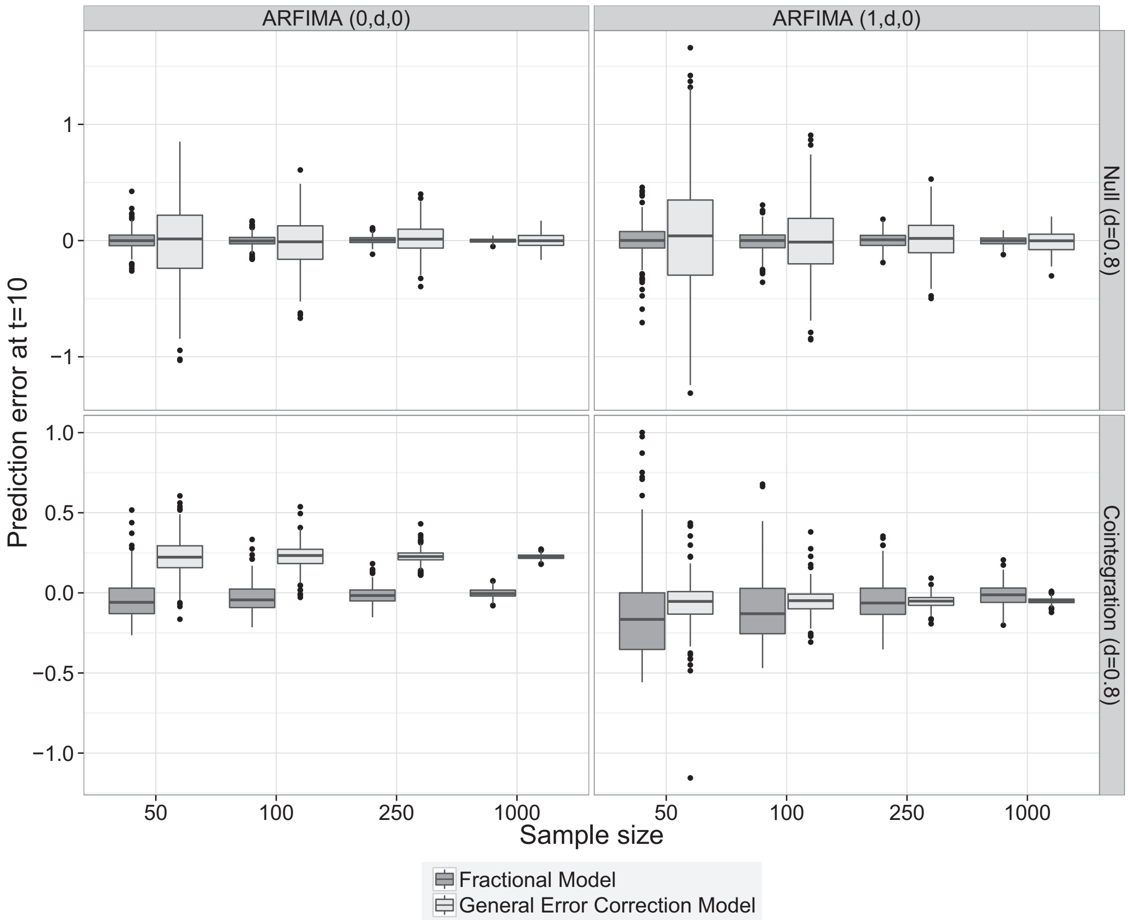

Fig. 3 Long-run effect of x on y at $t = 10$ when $d_x = d_y = 0.8$.
Notes. Results for long-run effect of x on y at $t = 10$ for the 16 simulation conditions in which $d_x = d_y = 0.8$ (Equation 9). The setup is based on the scheme described in Fig. 1.

both models get closer to the true effect as the sample size increases. This is not the case when there is a cointegrating relationship between the variables (lower-left panel). Although the prediction error for the two models is similar for $T = 50$, the prediction error for the FIM improves considerably as the sample size increases and approaches the true cumulative effect, $\sum_{t=1}^{10} \Delta y_t = 1.4$. However, the GECM consistently underpredicts the cumulative effect of x on y. This is to be expected since the GECM fails to account for the long-term persistence of the two variables.

Turning our attention to the right column, we see that the inclusion of short-run dynamics has the expected effect on the prediction error. When there is no relationship between the variables (upper-right panel), the prediction error increases (the boxes on the boxplots become larger), although it continues to be centered on the true cumulative effect, 0. However, when there is a cointegrating relationship (lower-right panel), both models perform poorly at small sample sizes. As was the case when no short-run dynamics were included, the FIM approaches the true cumulative effect, $\sum_{t=1}^{10} \Delta y_t = 2.7$, when the sample size increases, while the GECM provides ever more precise predictions of the wrong effect.

The results for the effect of x on y at $t = 10$, when $d_x = d_y = 0.8$, are shown in Fig. 3.[5] When there is no relationship between the two variables and no short-run dynamics (upper-left panel), the

[5]Note that unlike Fig. 2 this figure reports the effect of a one-unit shock to x at $t = 1$ on the value of y at $t = 10$, that is, not the cumulative effect on y from $t = 1$ to $t = 10$.

predictions from both models center on the true effect, $y_{t+10} = 0$, although the prediction error for the FIM is considerably smaller than the GECM. Estimates from both models get more precise as the sample size increases, as is to be expected.

However, when there is a cointegrating relationship between the two variables (lower-left panel), the two models provide diverging predictions. As in Fig. 2, the FIM provides predictions closer to the true effect, $y_{t+10} = 0.27$, which get progressively more accurate as the sample size increases. Predictions from the GECM, however, overpredict the value of y ten periods into the future, and provide ever more precise estimates of the wrong effect as the sample size increases. This is to be expected, since the GECM treats the data as $I(d=1)$, rather than $I(d=0.8)$, and, thus, overestimates the persistence of shocks to the two variables.

The right column of Fig. 3 shows that the inclusion of short-run dynamics increases the average prediction error considerably. When there is no relationship between the two variables (upper-right panel), the variability in predictions increases, although they remain centered on the true effect, $y_{t+10} = 0$. However, when there is a cointegrating relationship between the two variables (lower-right panel), the GECM outperforms the FIM at small sample sizes, while the FIM provides progressively better predictions and converges to the true effect $y_{t+10} = 0.56$, when $T = 1000$.

4 Conclusion

How should practitioners working with short time series that are potentially fractionally integrated proceed? The current simulation study contributes to the dialogue between Grant and Lebo (2016) and Keele, Linn, and Webb (2016) by comparing the performance of the GECM to fractional integration models (FIMs) for different sample sizes, different relationships between x and y, and with or without short-run dynamics. Several recommendations can be made based on the study.

First, the two methods provide similar estimates of the short-run effect of x on y across most of the simulation conditions. Estimates from both the FIM and the GECM have similar absolute bias and become less biased as the sample size increases, and the 95% coverage probability in most of the conditions is close to the desired 95% value. If primary interest is in the immediate effect of x on y, analysts should be able to use either method to recover the true effect. This is perhaps unsurprising, yet helpful to show as part of this discussion, since the presence of fractional integration primarily affects the persistence of shocks to a series over time, rather than the immediate effect itself.

Second, when there are no short-run dynamics (and the analyst assumes so *a priori*), the FIM performs equally well or outperforms the GECM over each of the sixteen simulation conditions with no short-run dynamics. While both models provide unbiased predictions when there is no relationship between the two variables, the predictions from the FIM are more precise, as they cluster more closely around the true effect for all models. However, when there is a cointegrating relationship between the two variables, predictions from the FIM are never worse than predictions from the GECM and they become progressively more accurate as the sample size increases. Thus, if primary interest is in the long-run effect of x on y and there are negligible short-run dynamics, analysts can use fractional integration methods, importantly, even for samples as small as $T = 50$.

Third, when relatively mild short-run dynamics are present (and the analyst does not know the order of short-run dynamics), the models perform differently based on the sample size. For small samples ($T = 50$ when interest is in the cumulative effect of x on y at $t = 10$ and $T = 50$ or $T = 100$ when interest is in the effect of x on y at $t = 10$), the GECM provides more accurate predictions than fractional integration methods and is less prone to produce wildly inaccurate predictions. Meanwhile, when the sample size increases, the FIM provides progressively better predictions. Thus, if primary interest is in the long-run cumulative effect of x on y and there are (unknown) short-run dynamics, analysts should only use fractional integration methods for samples with more than 250 observations.[6]

[6]Note that the short-run dynamics included are fairly mild and I expect the FIM to perform even worse if there are more complex dynamics at play—accordingly, the sample size requirements should be greater in such cases.

Collectively, these results suggest that the presence or not of short-run dynamics in political time series is critical for whether analysts should use fractional integration methods in short time series or apply simpler methods. If Grant and Lebo's (2016, 25) claim, that most time series used in political science are well described by a simple (0,d,0), is empirically accurate, the prospects for fractional integration methods in short time series are quite positive. However, if more complex short-run dynamics are at play, analysts should proceed carefully when they have short time series and provide estimated effects based on different assumptions about the underlying data-generating process. Any inferences made from such an analysis must be limited, as suggested by Keele, Linn, and Webb (2016, 28). Finally, when practitioners have relatively long time series ($T > 250$), there are clear benefits to using fractional integration methods to allow for the possibility that the underlying data-generating process is fractionally integrated.

Conflict of interest statement. None declared.

References

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

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

Byers, David, James Davidson, and David Peel. 2000. The dynamics of aggregate political popularity: Evidence from eight countries. *Electoral Studies* 19:49–62.

Carsey, Thomas M., and Jeffrey J. Harden. 2013. *Monte Carlo Simulation and Resampling Methods for Social Science*. Thousand Oaks, CA: Sage Publications.

Cowpertwait, Paul S. P., and Andrew V. Metcalfe. 2009. *Introductory Time Series with R*. New York: Springer.

De Boef, Suzanna, and Luke Keele. 2008. Taking time seriously. *American Journal of Political Science* 52:184–200.

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

Grant, Tayler, and Matthew J. Lebo. 2016. Error correction methods with political time series. *Political Analysis* 24:3–30.

Helgason, Agnar Freyr. 2016. Replication Data for: Fractional Integration Methods and Short Time Series: Evidence from a Simulation Study. http://dx.doi.org/10.7910/dvn/lvnze2, Harvard Dataverse.

Keele, Luke, Suzanna Linn, and Clayton Webb. 2016. Taking time with all due seriousness. *Political Analysis* 24:31–41.

Lebo, Matthew J., and Christopher Weber. 2015. An effective approach to the repeated cross-sectional design. *American Journal of Political Science* 59:242–58.

Lebo, Matthew J., Robert W. Walker, and Harold D. Clarke. 2000. You must remember this: Dealing with long memory in political analyses. *Electoral Studies* 19:31–48.

Nielsen, Morten Ørregaard. 2004. Optimal residual-based tests for fractional cointegration and exchange rate dynamics. *Journal of Business & Economic Statistics* 22:331–45.

———. 2010. Nonparametric cointegration analysis of fractional systems with unknown integration orders. *Journal of Econometrics* 155:170–87.

Sowell, Fallaw. 1992. Modeling long-run behavior with the fractional ARIMA model. *Journal of Monetary Economics* 29:277–302.