Conditional Relationships in Dynamic Models*

Zach Warner†

September 17, 2016

Abstract

Many political science theories include variables that interact over time. Recent papers have tested such theories using autoregressive distributed lag and error correction models that include multiplicative interactions. However, the theory underlying such models is underdeveloped, causing difficulties with (1) knowing when OLS estimates are consistent, (2) writing models without opaque parameter restrictions, and (3) interpreting results. This paper provides theoretical and practical guidance to address these problems. First, I define the conditions under which scholars can ensure consistent estimates. Second, I introduce a general model that imposes no constraints on how conditional relationships unfold over time. Third, I develop a flexible approach for interpreting such models. I demonstrate the advantages of this framework with simulation evidence and an empirical application.

Word count: 8,703.

*I thank John Ahlquist for many helpful discussions throughout the life of the paper. I am grateful to Lisa Blaydes, Will Jennings, Peter John, Mark Kayser, Nathan Kelly, and Jana Morgan for generously sharing their data. For their comments and suggestions, I also thank Rikhil Bhavnani, Brian Gaines, Jude Hays, Noam Lupu, Melanie Manion, Jon Pevehouse, and Alex Tahk. Previous versions of this paper were presented at Wisconsin and the annual meeting of the Society for Political Methodology.

†Ph.D. Candidate, University of Wisconsin–Madison. email: zwarner@wisc.edu, web: http://www.zachwarner.net.

1 Introduction

Political scientists are increasingly developing theories where variables interact over time. With an abundance of large panel datasets to test such theories, many scholars are turning to dynamic models—which allow the effect of a covariate on the response to change over time—because they yield estimates of long-run effects. Several recent papers have extended such specifications to include multiplicative interaction terms, wherein the effect of one covariate depends on another. For example, Nooruddin and Simmons (2006) find that International Monetary Fund programs reduce social spending most sharply in democracies, while Kono (2008) considers the effect of democratization on trade liberalization, conditional on the wealth of the trading partners. Similar applications have contributed to a variety of literatures across the discipline.[1]

These models are often intuitive representations of complex systems of variables. Yet the theory underlying them is underdeveloped, even for workhorse specifications such as autoregressive distributed lag and error correction models (ADLs and ECMs, respectively). This shortcoming causes three important problems:

1. *Estimation.* A large literature describes how scholars can ensure that ADLs and ECMs are appropriate for their data through extensive "pretesting," particularly by examining stationarity and cointegration. However, there is no guidance on what tests an interaction among covariates must satisfy, and under what circumstances. As a result, scholars cannot be certain that ordinary least-squares (OLS) will produce consistent estimates.

2. *Specification.* Typically, each variable of interest in an ADL or ECM suffers

[1] Other examples include Aklin and Urpelainen (2013); Alexiadou (2015); Andersen and Ross (2014); Arnold and Carnes (2012); Blaydes and Kayser (2011); Carlin et al. (2015); Chang (2008); Cheon and Urpelainen (2015); Enns, Kelly, Morgan, Volscho, and Witko (2014); Escribà-Folch (2012); Hobolt and Klemmensen (2008); Jennings and John (2009); Jensen and Mortensen (2014); Kayser (2009); Kellam and Stein (2015); Keller and Kelly (2015); Kono and Montinola (2013, 2015); Kono et al. (2015); Lipsmeyer and Zhu (2011); Miller (2015); Morgan and Kelly (2013); Ramirez (2013); Sorens (2011); Swank (2006); Tenorio (2014); Werner and Coleman (2015); Wright (2015). See also dynsim, a recent package for R and Stata that provides a user-friendly resource for studying dynamic models with interactions (Gandrud et al. 2016).

from some amount of autocorrelation: its present value is a function of its past values. Choosing which lagged values to include in the model generates different statements about how the researcher believes covariates affect the response over time. However, it is unclear what lags of an interaction represent, since either component variable may be lagged, or both. This complication may lead to specifications that imply dynamic effects that differ from what the author intends.

3. *Interpretation.* ADLs and ECMs include lags and differences that complicate interpreting estimates. Interactions add to this challenge by altering the textbook equations for quantities such as long-run effects. Failure to account for these complications may lead to incorrect inferences.

This paper aims to help scholars avoid these problems, providing theoretical and practical guidance for studying conditional relationships in dynamic models.[2] First, I begin by defining the conditions under which OLS produces consistent estimates: even if two covariates are stationary, their interaction may not be, and so must be explicitly tested for stationarity. However, if a model includes cointegrated variables, the same model with an added interaction is always cointegrated. Scholars can therefore ensure that OLS is an appropriate estimator simply by extending standard diagnostic procedures.

Second, I show how to specify and interpret a general model with an interaction. Absent extremely strong *a priori* beliefs about the data-generating process (DGP), scholars should proceed from a specification that allows covariates to interact freely across time. I demonstrate that such a model includes all possible cross-period interactions; two variables, each with p lags, generates $(p+1)^2$ interaction terms that capture a single conditional relationship.[3]

Third, I provide a unified approach to calculating quantities of interest. The quantities most commonly encountered in the literature—including short- and long-run

[2] I use "conditional relationship" and "interaction" interchangeably throughout the rest of the paper to refer to a multiplicative interaction.

[3] This generalizes to cases where their lags differ, for which an interaction generates $(p_1+1)(p_2+1)$ terms.

effects, as well as various lag lengths—emerge as special cases of more general inferential tools. I also show how various parameter restrictions can then be imposed on the general model, and how such restrictions have drastic, often unintuitive, consequences for inferences about the dynamic system.

I then use Monte Carlo simulations to examine two important questions about the tradeoffs of this approach relative to standard practice. First, scholars may be concerned that estimating a model with an interaction, and especially a general model with $(p + 1)^2$ interaction terms, will induce bias: adding parameters may drag the estimated error correction rate (ECR) downward, and can lead to overfitting when time series are short (Enns, Masaki, and Kelly 2014; Keele et al. 2016). Second, we do not know whether these risks outweigh the potential for bias that arises from invalid restrictions. The evidence presented here suggests that the general specification produces far more reliable inferences: while there is some evidence of overfitting, this risk is relatively minimal compared to the cost of an invalid restriction. My results suggest that when two variables do not interact over time, a general model is hardly worse than a restricted specification. But when they do, a general model recovers the true parameter values while a restricted specification can produce seriously biased estimates.

Finally, I illustrate how these findings contribute to more robust inferences with an empirical application.[4] Blaydes and Kayser (2011) present evidence that democracies redistribute the gains from economic growth to the poor more than do autocracies or hybrid regimes, as commonly assumed in formal theories of regime transition (e.g., Acemoğlu and Robinson 2006). I replicate and extend this study using the more general framework developed here. I find that the authors' conclusions rely on biased estimates and incorrect calculations of predicted effects: regime type has no discernible role in mediating the effects of growth. This evidence suggests that the link between regime type and inequality is weaker than is assumed in many political economy theories.

A key takeaway of this paper is that scholars should not unthinkingly add interaction terms to their dynamic models. Here I echo the chorus of findings that applied work

[4] The online appendix also includes replications of Morgan and Kelly (2013) and Jennings and John (2009), with similar results.

typically suffers from too little attention to data constraints, diagnostic tests, and specification decisions.[5] Like Grant and Lebo (2016) and Keele et al. (2016), I urge scholars to rigorously test the properties of their data at each stage of modeling and estimation. The goal of this paper is to provide a toolkit for scholars to maintain these best practices when extending dynamic models to include conditional relationships.

2 Estimating dynamic models

There are a great many ways of modeling dynamic dependence among variables. To keep the analysis tractable, I only study multiplicative interactions in ADLs and ECMs, since they are the most common conditional relationships in the most common time series models in political science (Box-Steffensmeier et al. 2014).[6]

Throughout the paper, I develop my argument with the ADL because it facilitates transparency. Nevertheless, my results also hold for the ECM since the two are simple linear reparameterizations of one another (Davidson and MacKinnon 1993; De Boef and Keele 2008).[7] The inferential framework I introduce in Section 3 demonstrates that any quantities of interest can be recovered from either model, and helps scholars do so (with derivations in the online appendix). Additionally, in the empirical application below, I work within the error-correction framework preferred by Blaydes and Kayser (2011), but provide the same findings from ADL estimates in the replication archive.

[5] For examples and discussion, see the contributions to a recent special issue of *Political Analysis*, particularly Esarey (2016); Freeman (2016); Grant and Lebo (2016); Helgason (2016); Keele et al. (2016).

[6] This scope condition rules out a number of alternative approaches. Most notably, many scholars have made use of models that directly estimate time-varying relationships, such as in a Kalman filter or Dynamic Conditional Correlations (DCC) framework (Beck 1989; Lebo and Box-Steffensmeier 2008). While such approaches are suitable for many applications, they are somewhat rare in political science, and so I leave discussion of their relative merits for studying conditional relationships to future research. For a monograph-length treatment of time-varying approaches, see Petris et al. (2009).

[7] This equivalence has recently been brought under scrutiny by Grant and Lebo (2016), who argue that (compared to the ADL) the ECM directly estimates different quantities, finds significant results more routinely, and thus makes mistaken inferences more likely. Yet this misdiagnoses the problem. The ECM leads to inferential errors not because of the parameterization of the model *per se*, but rather because scholars pay too little attention to appropriately deriving and interpreting quantities of interest. See Keele et al. (2016) for simulation evidence on their equivalence, and Enns et al. (2016) for further discussion.

I begin by writing the ADL in the general form:

$$y_t = \alpha_0 + \sum_{f=1}^{p} \alpha_f y_{t-f} + \sum_{g=1}^{q} \sum_{h=0}^{r} \beta_{g,h} x_{g,t-h} + \epsilon_t. \tag{1}$$

This specification is ADL(p,q;r), with p lags of y_t, r lags of $x_{g,t}$, and q covariates. The standard method of deciding on p and r is the Box-Jenkins methodology, an inductive modeling technique by which researchers account for temporal dependency within each variable and between variables in the dynamic system (Box and Jenkins 1970; Box-Steffensmeier et al. 2014). The goal is to ensure that ϵ_t is "white noise," with $\mathbb{E}(\epsilon_t,x_{g,t-h})= 0 \; \forall \; t, g, h$.

If this condition holds, OLS produces consistent estimates in two cases. The first case requires that all variables in the model are stationary.[8] Intuitively, we want to make general inferences about the relationship among these variables, but all of our data are sampled from a specific time window. In order for these inferences to be valid, the data must be representative of each process beyond the limits of the sample. Stationarity is sufficient to guarantee that this is the case, since it requires a variable to have constant mean, variance, and auto-covariance over time.[9]

If this condition does not hold, then OLS estimates will still be consistent so long as the dynamic system is cointegrated (Engle and Granger 1987; Granger 1986). Cointegration occurs when the combination of nonstationary variables produces a stationary variable—for instance, in Equation 1, when the xs and y generate a stationary ϵ.

All ADLs and ECMs must satisfy an additional requirement: equation balance (Grant and Lebo 2016). All time series in a model must have the same order of integration, or "length of memory." Keele et al. (2016: 299) explain: "Stable/stochastically bounded variables cannot cause (or be caused by) the path of a stochastically unbounded

[8] Throughout this paper I refer only to covariance (or weak) stationarity. See the online appendix for formal definitions and analysis for AR(1) and AR(p) cases. Note that nonstationarity occurs when any (not necessarily all) of the requirements for stationarity are violated.

[9] Constant auto-covariance is when the covariance between any two points in time is equal, depending only on the magnitude of the time lag ℓ between them, and not the particular periods for which it is computed.

variable. The time series must eventually diverge by larger and larger amounts." In such cases where variables are of differing orders of integration, each must be appropriately adjusted (usually differenced) prior to estimation to establish equation balance.

How can scholars ensure that a model with a conditional relationship satisfies the requirements of either case? The time-series properties of interaction terms are poorly understood, not least because it seems intuitive that the "memory" of xz would directly follow the properties of x and z. This has two problematic implications. First, the standard practice of diagnosing the time-series properties of all lower-order terms does not guarantee consistent OLS estimates. Second, and perhaps more importantly, it is unclear what properties of the interaction need to be tested, and how to do so. To alleviate these problems, I provide theoretical results establishing when OLS estimates are consistent for a model with an interaction and how to test for those conditions. Proofs are provided in the online appendix.

I begin with the first case, in which all variables are stationary. By definition, this case requires that interaction terms are also themselves stationary. Proposition 1 specifies where this holds.

Proposition 1. Assume x and z are covariance-stationary autoregressive stochastic series. A series xz composed of their multiplicative interaction is itself stationary if and only if $\mathbb{C}[\hat{x}, \hat{z}] = \mathbb{C}[\tilde{x}, \tilde{z}] \; \forall \; \hat{x}$ and $\tilde{x} \in \{x_t, x_{t+\ell}, x_t x_{t+\ell}\}$, $\hat{z}$ and $\tilde{z} \in \{z_t, z_{t+\ell}, z_t z_{t+\ell}\}$, where $t, \ell \in \mathbb{Z}_{\geq 0}$, and $\hat{x}$ ($\hat{z}$) and $\tilde{x}$ ($\tilde{z}$) differ only in t.

Intuitively, Proposition 1 says that even if two covariates are not functions of time, the manner in which they interact may be. This would violate stationarity, in which case OLS may not produce consistent estimates. To ensure that stationarity holds, all covariances between the interacted variables must be independent of time.

Practically, Proposition 1 suggests that scholars always need to explicitly test interaction terms for stationarity. This result can be made more intuitive by considering the only case in which stationarity is *guaranteed* to hold: when the two variables are stochastically independent.

Corollary 1. Assume x and z are covariance-stationary autoregressive stochastic series. If x and z are stochastically independent, a series xz composed of their multiplicative interaction is itself covariance-stationary.

Scholars modeling conditional relationships are doing so precisely because they believe that the two variables have a dependent relationship; if they are thought to be truly independent, then there should be no point in studying their interaction. Corollary 1 should therefore never hold in practice, and scholars cannot assume that the dependence itself is not a function of time. Only by pretesting interaction terms directly can we be sure that they are stationary.

An example helps illustrate how this problem might arise in practice. Suppose two variables are known to be stationary, but their correlation is a function of time, as in a Dynamic Conditional Correlations (DCC) framework. Such relationships are common, for instance, with variables such as economic growth and presidential approval (Lebo and Box-Steffensmeier 2008). Although each variable would pass stationarity tests individually, their interaction may be nonstationary, and OLS estimates for a model that includes this interaction may be inconsistent. This scenario is a straightforward example of the spurious regression problem, a well-understood and widely documented phenomenon (Grant and Lebo 2016; Yule 1926).

There are two ways to read Proposition 1. On one hand, this result adds to the already-substantial burden on researchers to pretest their data. It also implies that equation balance is much harder to achieve for a model with a conditional relationship, since the interaction's order of integration need not be the same as the component variables'. This finding suggests that only very rarely will scholars be able to ensure that the ADL or ECM is appropriate for their data. On the other hand, however, pretesting each variable for stationarity is standard practice in time series modeling. There are a number of tests that have been developed and implemented across statistical software for this purpose, including the ADF, Variance Ratio, Modified Rescaled Range, and KPSS tests and variants thereof (for a recent review, see Box-Steffensmeier et al. 2014). These same tests can be used for interaction terms. Scholars should take extra caution when estimating an ADL or ECM specification with a conditional relationship, but the diagnostic procedure remains the same.[10]

Results are perhaps more encouraging for the second case, in which the variables are cointegrated.

Proposition 2. If a cointegrating vector exists for a dynamic system composed of a response y and two covariates x and z, then a cointegrating vector also exists for a system including the multiplicative interaction xz.

This Proposition states that if the variables in the model are cointegrated, then adding an interaction term makes no further demands of the data, provided the equation remains balanced. OLS estimates will still be consistent. Moreover, this condition is only sufficient, not necessary: it is possible that the dynamic system *becomes* cointegrated when an interaction is included. Scholars should therefore explicitly test models with a conditional relationship for cointegration. As Enns, Masaki, and Kelly (2014) argue, such tests should be conducted for all ADLs and ECMs, and a number of diagnostics already exist.[11] Thus, adding a conditional relationship imposes no new hardships on the researcher. Together, Propositions 1 and 2 establish that ADLs and ECMs are appropriate models of conditional relationships in dynamic systems, so long as scholars extend the standard diagnostic procedures to include interaction terms.

These results speak only to consistency, not the other desirable properties, of OLS estimates. Yet there are at least two ways in which interaction terms can lead to bias. First, as Enns, Masaki, and Kelly (2014) note, increasing the number of covariates exacerbates bias for the parameter relating to the lagged response, the error-correction rate (ECR). Since the ECR is present in all calculations of dynamic effects, adding interactions may threaten inferences about long-term relationships among variables. The second problem is overfitting. There is much less information in time series data

[10] In addition to stationarity and unit-root tests, Grant and Lebo (2016) suggest fractional integration methods (FIM) to pretest data that may be neither stationary nor integrated. However, Keele et al. (2016) argue that the FIM approach requires much longer time series than are typically available in order to reliably estimate orders of integration. Simulation evidence indicates that the standard ADL/ECM without FIM performs well under a variety of circumstances (Esarey 2016; Helgason 2016). Whichever pretesting approach scholars decide is suitable for their application, my results indicate that it must be extended to conditional relationships.

[11] In particular, see the Engle and Granger (1987) and Johansen (1988) tests.

than sample sizes suggest, making it likely that scholars who add extra covariates are simply modeling noise. Results from such specifications will be unstable (Keele et al. 2016). ECR bias and overfitting are important problems that scholars need to grapple with as they study dynamic models, and there is no clear consensus on how to avoid them, other than to keep models simple.

I take up these concerns in the Monte Carlo analysis below. To preview my results, these simulations suggest that neither ECR bias nor overfitting alone should discourage researchers from estimating interactions, as the potential costs appear relatively slight. Even still, in general, I urge scholars to carefully consider these tradeoffs before including interactions—or any new covariates—in a dynamic model.

3 Studying conditional relationships

Despite a large literature on the properties of ADLs and ECMs, scarce attention is paid to specifying relationships of interest. This lacuna is unfortunate, as even small variations among relatively simple models can imply very different dynamics. In practice, scholars are unable to ensure that the relationships implied in their statistical models match those of their theory, nor can they ensure that they are using appropriate formulas for drawing inferences from their models. These problems generate incorrect or incomplete interpretations.

In this section I introduce the general model that interacts two variables, each autoregressive of order 1, or AR(1). This model allows a covariate interaction to unfold freely across time. I then provide a flexible framework for drawing inferences from this model, which can be used to study any ADL or ECM,[12] and discuss the use of parametric bootstrapping for quantifying uncertainty. Finally, I note a few potential parameter restrictions that can be imposed, and how they affect quantities of interest.

[12] See the online appendix for full derivations of these quantities.

3.1 The general model

The most general ADL that captures a conditional relationship between two AR(1) variables is given by

$$
\begin{aligned}
y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 z_t + \beta_3 z_{t-1} + \\
\beta_4 x_t z_t + \beta_5 x_{t-1} z_t + \beta_6 x_t z_{t-1} + \beta_7 x_{t-1} z_{t-1} + \epsilon_t,
\end{aligned} \tag{2}
$$

which translates into an equivalent ECM,

$$
\begin{aligned}
\Delta y_t = \gamma_0 + \gamma_1 y_{t-1} + \theta_0 \Delta x_t + \theta_1 x_{t-1} + \theta_2 \Delta z_t + \theta_3 z_{t-1} + \\
\theta_4 \Delta x_t z_{t-1} + \theta_5 x_{t-1} \Delta z_t + \theta_6 \Delta x_t \Delta z_t + \theta_7 x_{t-1} z_{t-1} + \epsilon_t.
\end{aligned} \tag{3}
$$

Since x and z are AR(1), both models include two periods for each variable. Specifying a conditional relationship therefore requires 2×2 interaction terms in addition to the lower-order variables. More generally, interacting two AR(p) processes generates $(p+1)^2$ new parameters, which allows the covariates to interact across any points in time. These terms all appear in calculating quantities of interest; eliding any implies a restriction on the dynamic system.

This specification is a reasonable starting point for most applications. First, the cost of invalid parameter restrictions is enormous: biased estimates and incorrect inferences (De Boef and Keele 2008). Erroneously constraining any of the β or θ terms in the general model to zero breaks the link between different periods of the same variable. Failing to account for the "memory" of each process in this way induces bias for not only the covariate directly constrained, but also among any interacted variables and lower-order terms. For instance, setting $\beta_4 = 0$ when the true DGP is Equation 2 biases estimates of all xz terms, and through the conditional relationship, all x and z terms. The Monte Carlo exercise below underscores the magnitude of these problems: in some cases, invalid restrictions can produce essentially random estimates.

Second, the general model is attractive because it privileges information in the data over broad theoretical intuition. Even the most precise theories are generally silent on exactly *when* covariates interact, so a conservative approach is to allow for all possibilities. Moreover, where theory does provide such strong guidance on the timing of the interaction, these hypotheses can be empirically tested by estimating the general model. Rather than impose restrictions based on *a priori* beliefs, this approach allows scholars to treat parameter estimates as evidence of how the dynamic system behaves. Finally, this model allows scholars to parse different components of a single conditional relationship. The interaction xz can vary because of a shift in x alone, z alone, or both. By estimating all cross-period interactions, scholars can study the effect of each change on the dependent variable separately. All of these features contribute to a richer set of inferences.

3.2 A framework for inference

Any model is only as useful as the information it provides. To help get the most from the general model in Equation 2, I develop a flexible approach to interpreting parameter estimates. The intuition of this framework is that scholars are typically interested in change in the response y as a function of change in (or "shock" to) a variable x, which implies that anything we want to know about the dynamic system can be found just by manipulating derivatives. This method yields three types of quantities that can be studied in a dynamic setting: period-specific effects, cumulative effects, and threshold effects. While I present results for a system in which all variables are AR(1), each quantity easily extends to AR(p) processes. And although the results are expressed in ADL parameters, the equivalent quantities for the ECM are provided in the Appendix.

3.2.1 Period-specific effects

Period-specific effects are defined as the change in y at any time as a result of a shock to x at time t. In other words, they answer: "what will happen to the response at some future date if a covariate shock occurs today?" For any period $t + j$, where t and j are integers, this quantity can be calculated by specifying the model for y_{t+j} and differentiating with respect to x_t. From Equation 2, note that

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 y_{t-1} + \ldots, \\ y_{t+1} &= \alpha_0 + \alpha_1 y_t + \ldots, \\ &\vdots \\ y_{t+j-1} &= \alpha_0 + \alpha_1 y_{t+j-2} + \ldots, \\ y_{t+j} &= \alpha_0 + \alpha_1 y_{t+j-1} + \ldots, \end{aligned}$$

so that each expression can be substituted for the next period's lagged y. Thus, we can start in period $t + j$ and recursively substitute previous periods of y until the right side of Equation 2 appears. Differentiating with respect to x_t therefore yields the general expression for period-specific effects:

$$\frac{\partial y_{t+j}}{\partial x_t} = \begin{cases} 0 & \text{for } j \in \mathbb{Z}_{<0}, \\ \beta_0 + \beta_4 z_t + \beta_6 z_{t-1} & \text{for } j = 0, \\ \alpha_1^j \left(\beta_0 + \beta_4 z_t + \beta_6 z_{t-1}\right) + \alpha_1^{j-1} \left(\beta_1 + \beta_5 z_{t+1} + \beta_7 z_t\right) & \text{for } j \in \mathbb{Z}_{>0}. \end{cases} \tag{4}$$

This is an intuitive result: the effect of the shock is transmitted through the coefficient on each x_t term, decaying at a rate determined by the memory of y—the error correction rate, α_1. Since x is itself autoregressive, the shock also decays through its own lagged values, weighted by exactly α_1 less. All of these parameters are directly estimated, so any period-specific effect is straightforward to calculate from the model output.

Scholars typically restrict attention to the case where $j = 0$, which De Boef and Keele (2008) refer to as the "short-run effect." Yet this quantity—better understood as the *instantaneous* effect of x on y—is only one among many period-specific effects that may be useful for studying a dynamic system. For example, Casillas et al. (2011) examine the effect of public opinion on Supreme Court decisions. They conclude that for every 1% liberal shift in public mood, the Court increases liberal reversals by 1.05% in the long run, and that 97% of this effect occurs after the first two periods. However, calculating period-specific effects reveals a different understanding of this relationship. The instantaneous effect at time t is 1.59, the effect in period $t+1$ is -0.42, in period $t+2$ it is -0.17, and so on. Thus, rather than monotonically increasing over time, the effect of public opinion on judicial decisions is immediately large but diminishes significantly. Period-specific effects can help clarify such relationships, improving our understanding of dynamic systems.

3.2.2 Cumulative effects

Of particular interest to scholars estimating dynamic models are cumulative effects, which provide information about relationships over time. These can be calculated by summing period-specific effects: for any time window $[t+h, t+k]$, the total change in y as the result of a shock to x is

$$\sum_{j=h}^{k} \frac{\partial y_{t+j}}{\partial x_t} \equiv \frac{\partial y_{t+h}}{\partial x_t} + \frac{\partial y_{t+h+1}}{\partial x_t} + \cdots + \frac{\partial y_{t+k}}{\partial x_t},$$

where h and k are non-negative integers and $h < k$. Note that each derivative is with respect to x_t; this sum is agnostic about variation in x following an initial shock. This expression can be solved by substituting from Equation 4 and simplifying the resulting geometric series.

Recall from Equation 4 that the instantaneous effect of x on y is slightly different from all other period-specific effects, as the shock to x has not yet worked its way through the system of variables. Unsurprisingly, this special case also shows up in cumulative effects. For $h > 0$, the expression for the cumulative effect is given by

$$\sum_{j=h}^{k} \frac{\partial y_{t+j}}{\partial x_t} = \frac{(\beta_0 + \beta_4 z_t + \beta_6 z_{t-1})\left(\alpha_1^h - \alpha_1^{k+1}\right) + (\beta_1 + \beta_5 z_{t+1} + \beta_7 z_t)\left(\alpha_1^{h-1} - \alpha_1^k\right)}{1 - \alpha_1}, \quad (5)$$

but for $h = 0$, it is simply

$$\sum_{j=0}^{k} \frac{\partial y_{t+j}}{\partial x_t} = \frac{(\beta_0 + \beta_4 z_t + \beta_6 z_{t-1})\left(1 - \alpha_1^{k+1}\right) + (\beta_1 + \beta_5 z_{t+1} + \beta_7 z_t)\left(1 - \alpha_1^k\right)}{1 - \alpha_1}. \quad (6)$$

As with period-specific effects, there is one special case of a cumulative effect that receives the majority of scholarly attention: the total change in y resulting from a change in x. This quantity is known as the long-run effect (LRE) or long-run multiplier (LRM), and can be found by setting $h = 0$ and allowing k to approach infinity:

$$\sum_{j=0}^{\infty} \frac{\partial y_{t+j}}{\partial x_t} = \frac{\beta_0 + \beta_1 + \beta_4 z_t + \beta_5 z_{t+1} + \beta_6 z_{t-1} + \beta_7 z_t}{1 - \alpha_1}. \tag{7}$$

This derivation demonstrates that the LRE is more precisely the *total* cumulative effect, since it maximizes the window over which change in y is summed. By no means is the LRE the only quantity available to scholars for describing long-run relationships. There are many time intervals, short and long, that can be constructed to study the behavior of the dynamic system.

These other cumulative effects deserve greater attention in the literature for three reasons. First, extrapolating from the LRE to finite-period relationships is unreliable. The difference between the cumulative effect from zero to k and from zero to infinity is often non-trivial.[13] It is also increasing in α_1, so the LRE is less useful for making inferences about finite periods where y is more strongly autoregressive. In other words, the longer it takes for effects to accumulate or dissipate over time, the worse the end result describes what's going on in the interim. The amount by which the LRE differs from a finite-period cumulative effect is also increasing in the LRE itself. Thus, even very large infinite-period relationships are not necessarily good approximations of finite-period relationships.

Second, many scholars work under data constraints that make interpreting the LRE difficult to justify. For example, if the output of a dynamic model indicates that the response is slow to adjust to a covariate shock, but the panel only covers a short time series, then such inferences may be more conjecture than calculation. Statements about cumulative effects over a few periods ameliorate the problem of extrapolating from short time series.

[13] This difference can be found by subtracting Equation 6 from Equation 7, and equals $\frac{\alpha_1^k(\alpha_1[\beta_0+\beta_4 z_t+\beta_6 z_{t-1}]+\beta_1+\beta_5 z_{t+1}+\beta_7 z_t)}{1-\alpha_1}$.

Finally, finite-period cumulative effects deserve greater attention because they are arguably more interesting and substantively important. For instance, the question "What is the effect of an economic shock on public opinion over the length of a presidential term?" cannot meaningfully be answered with the LRE. Even if it is possible to calculate how economic downturn affects presidential popularity from now until forever, the many elections in between may make this exercise more of a curiosity than an empirically relevant prediction. Examining cumulative effects over various time windows allows scholars to make more precise statements about the substantive significance of their findings.

3.2.3 Threshold effects

Threshold effects are defined as the number of periods following a shock to x required for the total change in y to reach some theoretically-relevant value λ. These are closely related to the lag lengths discussed in De Boef and Keele (2008), which define λ as a proportion (δ) of the LRE. Thus threshold effects may answer "How many years would it take for democratization to raise per capita income by $\$100$?" while lag lengths may answer "How many years after democratization does 80% of total change in income accrue?"

Threshold effects and lag lengths are calculated by setting Equation 6 equal to some λ and solving for the period k. How to interpret them depends on whether the cumulative effect of x on y is monotonically increasing over time. If it is, then only thresholds $|\lambda| < |\text{LRE}|$ produce sensible answers, since a threshold bigger than the total effect will never be reached. If it is not, then we choose a threshold $|\text{LRE}| < |\lambda|$ and solve for the period k during which the cumulative effect drops below λ in absolute value. Since thresholds are difficult to compute in closed analytic form, we can simply calculate cumulative effects over increasing k and visually inspect the results to find the period where the predicted values cross λ. Evaluating cumulative effects over a variety of periods is an important part of interpreting such models in general, so thresholds can be studied with trivial added effort.

Scholars commonly discuss these effects only in the context of median and mean lag lengths, which describe the period at which half ($\delta = .5$) and all ($\delta = 1$) of the LRE has accumulated, respectively. De Boef and Keele (2008: 192) argue that scholars should always report these "forgotten quantities." Yet the lag length for $\delta = 1$ is typically either nonsensical or uninteresting. If the cumulative effect of x on y is monotonic, then $\delta = 1$ is undefined, since the LRE is always greater (in absolute value) than any finite-period cumulative effect to some decimal place. If it is non-monotonic, then $\delta = 1$ simply yields $t = 0$, as in the Casillas et al. (2011) example above. The inferential value of this quantity is therefore limited.

More generally, thresholds of interest need not be set to one of only two values, but rather can be used to examine any quantile of the LRE. Nor must threshold effects be expressed as a proportion of the LRE at all. Absolute levels of change in y may be of greater interest, especially for policy applications. As with the other quantities described above, a variety of threshold effects can be used to convey information about the specific dynamic system under examination.

3.2.4 Uncertainty

Period-specific, cumulative, and threshold effects can all be reported with statements of uncertainty. Perhaps the most easily implemented method of calculating uncertainty is parametric bootstrapping, which requires fitting a model and then sampling from the distribution of the estimated parameters. Building on King et al. (2000), we can describe these steps for a maximum likelihood framework:

1. Estimate the model, e.g. Equation 2, from which we wish to draw inferences.
2. Store the point estimates ($\hat{\boldsymbol{\beta}}$) and variance-covariance matrix ($\mathbb{C}(\hat{\boldsymbol{\beta}})$).
3. Take M draws from the multivariate normal $\mathcal{N}_{MV}\left(\hat{\boldsymbol{\beta}}, \mathbb{C}(\hat{\boldsymbol{\beta}})\right)$ and save the output.
4. Define S scenarios of interest to study. For instance, we may hold all variables at their central tendency but vary one along its interquartile range by S increments. With dynamic models, these scenarios need to satisfy the constraint $\Delta y_t = y_t - y_{t-1}$ (and similarly for x and z) to be empirically relevant.

5. For each $s \in S$ and $m \in M$, calculate quantities of interest.

6. For each $s \in S$, take quantiles over the M draws, e.g. the .025$^{\text{th}}$ and .975$^{\text{th}}$, to approximate 95% confidence intervals.

This procedure yields statements of uncertainty for every quantity of interest for every scenario $s \in S$. Example code for this procedure is provided in the replication archive.

Parametric bootstrapping offers significant advantages over the standard practice for estimating uncertainty. Scholars generally rely on the delta method, but this approach is not easily extended to the more general quantities described above, nor to more complex models. In practice, this often leads to interpreting point estimates in the absence of uncertainty bounds, as De Boef and Keele (2008) do with lag lengths, or only interpreting instantaneous effects and the LRE. Parametric bootstrapping resolves this problem.

More than just protecting inferences, this approach allows scholars to get more from the data. For instance, suppose the instantaneous effect of a shock to a covariate is large. Cumulative effects may be statistically and substantively significant for a number of years, but they decay very slowly until the total effect it is indistinguishable from zero. For such cases, it would be true but misleading to conclude that there is no long-run relationship between the variables because the null cannot be rejected for the LRE. By giving uncertainty estimates for all of the quantities we want to study, parametric bootstrapping allows scholars to make much richer inferences about dynamic systems.

3.3 Restrictions on the general model

Scholars may wish to compare predicted effects from the general model with those from a restricted specification, in which one or more β terms are set to zero. Yet parameter restrictions often have drastic consequences for calculating the quantities outlined above. In general, formulas derived from one model do not match those from another model, even if only a single term differs between them. Choosing to interpret a restricted specification therefore requires that scholars derive quantities of interest *for that model*.

Kono and Montinola (2013) provide an example. They investigate the effect of official development assistance (ODA) on military spending, conditional on regime type. Ignoring control variables, their specification is the ECM

$$\Delta y_t = \gamma_0 + \gamma_1 y_{t-1} + \theta_0 \Delta x_t + \theta_1 x_{t-1} + \theta_2 \Delta z_t + \theta_3 z_{t-1} + \theta_4 \Delta x_t z_{t-1} + \theta_7 x_{t-1} z_{t-1} + \epsilon_t,$$

where y is military expenditure as a proportion of gross national income (GNI), x is ODA as a proportion of GNI, and z is regime type (variously measured). This specification restricts $\theta_5 = \theta_6 = 0$, implying that there is no interaction that involves change in regime type, consistent with the authors' theoretical expectations. Yet deriving the LRE for this model yields $\frac{\theta_1 + \theta_7 z_t - \theta_4 \Delta z_t}{-\gamma_1}$, indicating that the effect of a change in ODA is conditional on change in, not just level of, democracy.[14] In general, this disjuncture between standard LRE calculations and relationships implied in a restricted model could lead to incorrect inferences.

A restriction of particular interest to scholars is where one of the interacted variables is not a time series, but rather a vector of indicators (e.g., country fixed effects). Since this vector **z** is not stochastic and therefore not autoregressive, only one period of **z** needs to be interacted with each period of x to capture all cross-time interactions. Maintaining notation from above, the resulting model is given by

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 \mathbf{z} + \boldsymbol{\beta_4} x_t \mathbf{z} + \boldsymbol{\beta_5} x_{t-1} \mathbf{z} + \epsilon_t.$$

Quantities of interest can be derived from this specification using the same general framework.

4 Simulation evidence

The theoretical discussion above has raised two important questions. First, adding parameters to an ADL or ECM can lead to ECR bias and overfitting. How significant are these problems for the general model in Equation 2, with its $(p+1)^2$ interaction

[14] See Equation A2 for the LRE of an unrestricted ECM, from which this restricted LRE follows.

parameters? Second, how do these problems trade off against the potential for bias arising from invalid parameter restrictions in more parsimonious models? To address these questions, I simulate data from four DGPs:

$$y_t = \rho_y y_{t-1} + \eta_{y,t}, \tag{8}$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 z_t + \beta_3 z_{t-1} + \epsilon_t, \tag{9}$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 z_t + \beta_3 z_{t-1} + \beta_4 x_t z_t + \epsilon_t, \tag{10}$$

$$\begin{aligned} y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 z_t + \beta_3 z_{t-1} + \\ \beta_4 x_t z_t + \beta_5 x_{t-1} z_t + \beta_6 x_t z_{t-1} + \beta_7 x_{t-1} z_{t-1} + \epsilon_t, \end{aligned} \tag{11}$$

where $x_t = \rho_x x_{t-1} + \eta_{x,t}$ and $z_t = \rho_z z_{t-1} + \eta_{z,t}$. For simplicity, I refer to the DGPs in Equations 8-11 as the "spurious," "no interaction," "restricted," and "full interaction" DGPs, respectively. I hold the α and β terms fixed across all simulations. For each of the DGPs, I examine $3^4 = 81$ cases, varying the time series length $n \in \{50, 100, 500\}$ and the autoregressive parameters $\rho_x, \rho_z, \rho_y \in \{0.1, 0.5, 0.9\}$. For each of the 324 (4×81) experimental conditions, I create 500 datasets, with each variable's starting value and all errors $(\epsilon, \eta_x, \eta_z, \eta_y)$ drawn from $\mathcal{N}(0, 1)$, before estimating the models in Equations 9-11 and storing the results.

Turning first to the question of ECR bias, Figure 1 presents clear evidence that the general model does not produce worse estimates of α_1 than does a simpler model. The top panel plots estimates from the model without any interaction terms (in triangles), the restricted model with just $x_t z_t$ estimated (squares), and the general model (circles), for each of the four DGPs. Each point represents the mean estimate across 500 replicates under the same experimental condition, with dashed lines for the true values (each ρ_y in the "spurious" DGP, and $\alpha_1 = 0.5$ for the others). The general model provides estimates at least as close to the true value as those of a simpler model across all DGPs, and performs far better when the DGP includes a complex conditional relationship. This performance is reflected in Table 1, which presents absolute ECR bias ($|\hat{\alpha}_1 - \alpha_1|$ calculated for each replicate), averaged across all conditions under each DGP. Overall, the general model produces the lowest absolute bias.

The bottom panel of Figure 1 plots coverage probabilities for estimates of α_1—the

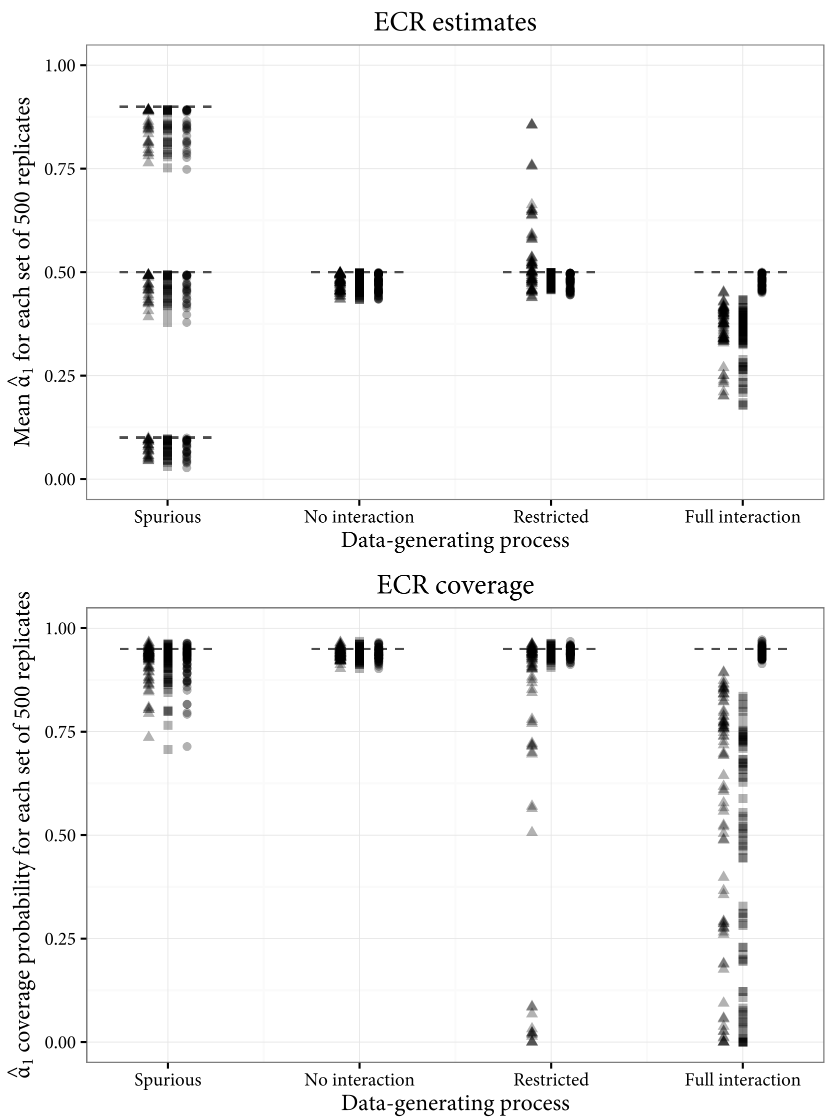

Figure 1: Monte Carlo simulation results for the error-correction rate, α_1. Triangles represent estimates from the model without an interaction, squares are from a restricted model, and circles are from the general model. Mean estimates are plotted against the true values (dashed lines) in the top panel. The bottom panel plots the proportion of 500 replicates under each condition for which 95% confidence intervals contain the true value.

Table 1: Average absolute bias, $\hat{\alpha}_1$

	DGP			
	"Spurious"	No int.	Restricted	Full
No int. model	0.28	0.06	0.10	0.14
Restricted model	0.28	0.06	0.05	0.16
General model	0.28	0.06	0.05	0.05

proportion of the 500 replicates under each condition for which the 95% confidence interval (constructed via parametric bootstrap) for $\hat{\alpha}_1$ contains the true value. Again, the evidence suggests that the general model recovers better estimates of the ECR, with higher coverage probabilities across the board than those of simpler specifications. Across all conditions, the mean coverage rate for the general model is 93%, very close to the 95% rate predicted by theory and much better than the 79% and 81% rates for the no interaction and restricted models, respectively. Even when we ignore the full interaction DGP—the most favorable case for the general model—the coverage rate (93%) remains as accurate as that of the restricted model (93%) and higher than the no interaction model (87%). Together, these data suggest that across a broad range of experimental conditions, estimating the general model does not exacerbate ECR bias.

Next, I evaluate whether there is evidence that the general model overfits. If overfitting is a problem, we should see (1) worse out-of-sample predictive power for the general model, relative to more parsimonious specifications, and (2) higher Monte Carlo variance (i.e., unstable coefficient estimates) across replicates under the same experimental condition.

Figure 2 plots the mean RMSE for out-of-sample predictions generated through 5-fold cross-validation.[15] With short time series and no interaction in the DGP, the general model produces RMSEs about 5-15% larger than those of simpler specifications. However, where there is an interactive process in the DGP, simpler specifications produce RMSEs approximately 25-75% larger than those of the general model—and this effect does not diminish as sample size increases. As a result, the grand mean

[15] To ease computational constraints, I take a random 20% sample from each set of replicates.

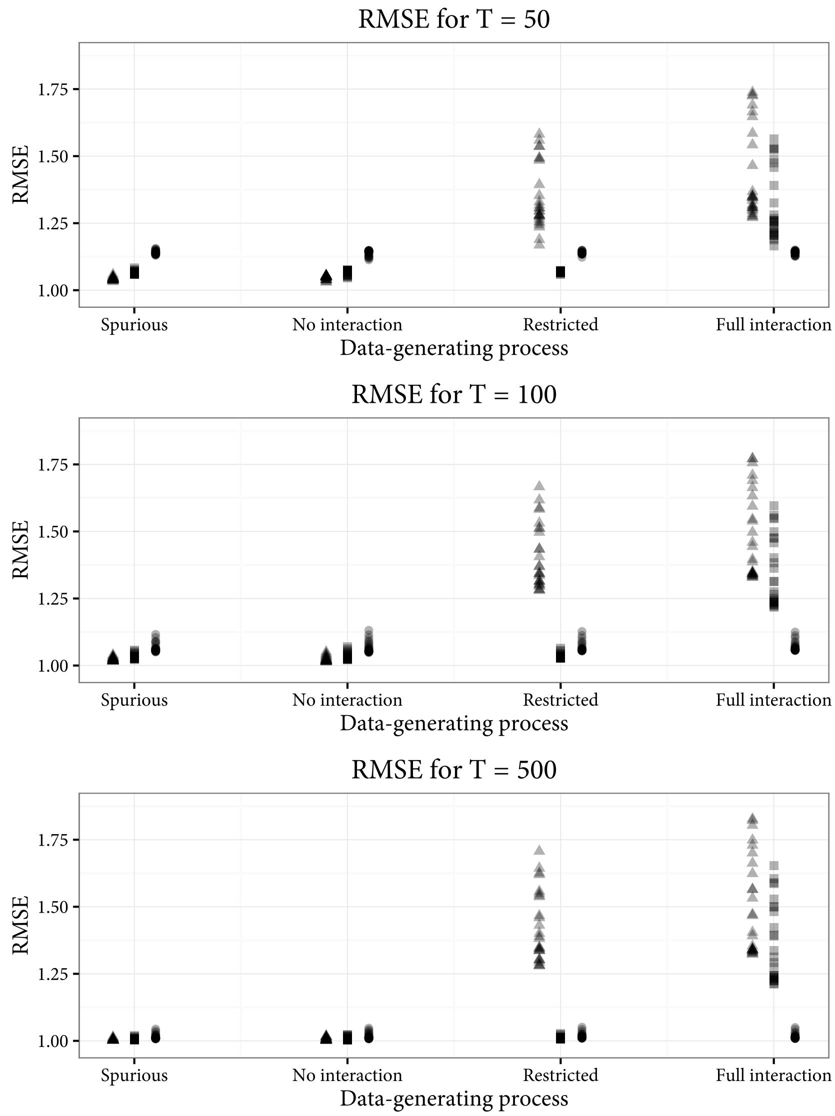

Figure 2: Monte Carlo simulation results for RMSE from 5-fold cross-validation. Triangles represent RMSE from the model without an interaction, squares are from a restricted model, and circles are from the general model. The top panel is for a time series of length $T = 50$, the middle is $T = 100$, and the bottom is $T = 500$. Each point represents the mean RMSE from a 20% random sample of each set of 500 replications under each experimental condition.

Table 2: Mean Monte Carlo variance across all simulations

	Mean Monte Carlo variance						
	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
No int. model	0.04	0.01	0.07	0.05	0.04	0.03	NA
Restricted model	0.03	0.01	0.02	0.04	0.02	0.02	0.01
General model	0.02	0.01	0.02	0.02	0.02	0.02	0.02

RMSE for the general model is 1.07, lower than that of the no interaction (1.23) and restricted (1.11) models.

To compare the stability of estimates across models, Table 2 reports the variance for each set of 500 replicates, averaged across all experimental conditions. The general model has the smallest variance for all parameters except the coefficient on $x_t z_t$, $\hat{\beta}_4$. This result is consistent with slight overfitting, but with a difference of 0.01, appears to be only a minor concern: the average 95% quantiles of $\hat{\beta}_4$ for each set of replicates has a width of 0.45 for the general model, compared to 0.38 for the simple interaction model. Together, the RMSE and variance evidence suggests that overfitting is a relatively minor concern for the general model.

I now turn to the second question: what are the potential costs of estimating a restricted model? Figures 1 and 2 demonstrate that invalid restrictions on an interactive process—including ignoring it altogether—generate substantial ECR bias, poor coverage rates, and poor predictions. To highlight just one result, inferences about the conditional relationship suffer drastically from invalid restrictions: Figure 3 plots estimates for the coefficient on $x_t z_t$ for the restricted and full models when the DGP includes a complex interaction. Mis-specification produces severe bias, with many 95% quantiles completely missing the true value, and highly unstable mean estimates.

In sum, the cost of estimating a general model are slight, in contrast to the problems that arise from invalid restrictions. These results should give scholars pause. To my knowledge, no political science study with a dynamic interaction has estimated the general specification, instead proceeding directly from a restricted model. My results indicate that this current standard operating procedure likely produces seriously biased

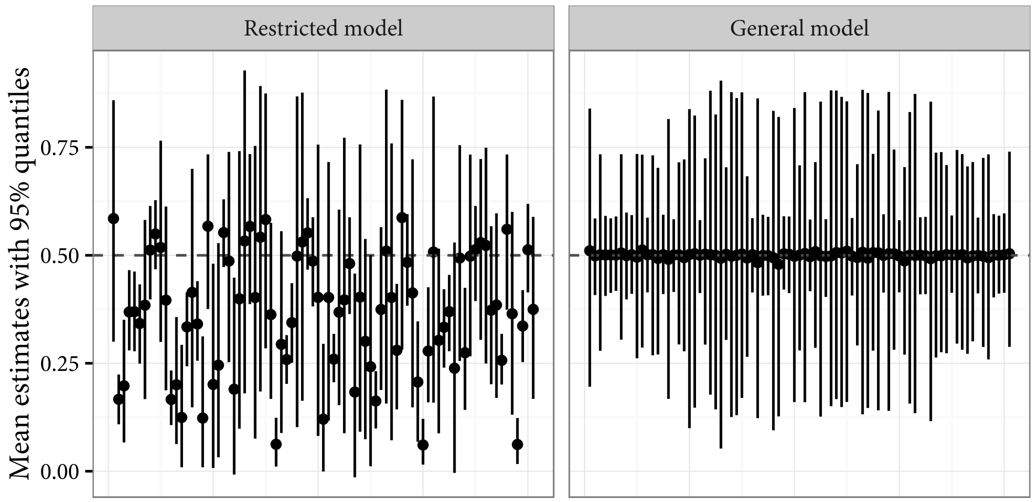

Figure 3: The consequences of invalid restrictions. Monte Carlo simulation results for β_4 when the DGP includes a complex conditional relationship. Points represent mean estimates from 500 replicates of each experimental condition, with lines for 95% quantiles.

estimates and unreliable inferences.

5 Application: democracy, growth, and inequality

In this section I demonstrate how the modeling and inferential framework developed above can improve our understanding of important dynamic relationships.[16] (Blaydes and Kayser 2011: "B & K") investigate a central problem in political economy: whether regime type conditions pro-poor growth in developing countries. Many formal models of regime transition rely on the assumption that democracies redistribute better than autocracies (Acemoğlu and Robinson 2006; Boix 2003), yet the evidence for this assumption is mixed (Houle 2009).

[16] This application uses a panel, which is less than ideal for illustration purposes. However, all of the political science applications of dynamic models with interactions—including all those in footnote 1—use either panel data or a very short time series, which poses equally serious challenges for inference. In the online appendix, I conduct a similar replication exercise for two further studies, Morgan and Kelly (2013) and Jennings and John (2009), the latter of which does not use a panel.

B & K make two innovations in addressing this question. First, they theorize that clientelism may be driving inequality reduction in democracies, generating a new empirical implication: if the gains from growth are shared through non-programmatic redistribution, then scholars should expect inequality reduction primarily in "hybrid" (or semi-democratic) regimes, where provision of clientelist goods is most common. Second, B & K introduce average daily caloric consumption as a proxy for inequality. Biological limits on the number of calories an individual can consume mean that the rich are not able to capture economic growth as it is transmitted through the production of basic foodstuffs, unlike increases in the value of capital or physical assets. Moreover, since the middle and upper classes consume near-ideal calories, any increase in the national average is likely to be driven by gains among the poor. This interpretation of consumption habits is substantiated by comparing calories with inequality figures generated elsewhere (Deininger and Squire 1996). Thus B & K hypothesize that the effect of growth on calories consumed is larger in hybrid regimes and democracies than in autocracies.

5.1 Data and estimation

Data on caloric consumption are gathered annually by the Food and Agriculture Organization, covering nearly every country in the world over the period 1961-2003. The figures are calculated by taking the initial stock of food, adding what was produced or imported, and subtracting what was exported, estimated to have been wasted, or left in the national stock at end of year. The result provides an estimate of annual consumption, which is then normalized by population. For the independent variables, B & K bin country-years by Polity IV regime type (Marshall et al. 2011), creating three categories: autocracies (≤ -7), hybrid regimes (-6 to 6), and democracies (≥ 7). Growth is calculated by first-differencing gross domestic product (GDP) per capita, which is expressed in hundreds of constant 2000 US dollars. As this literature is concerned primarily with developing countries, Blaydes and Kayser (2011) drop country-years with a GDP per capita of $\$10,000$ or higher. I follow their coding decisions to ensure comparability.[17]

I first replicate B & K's preferred model, given by

$$\Delta y_t = \gamma_0 + \gamma_1 y_{t-1} + \theta_0 \Delta x_t + \theta_1 x_{t-1} + \boldsymbol{\theta}_3 \mathbf{z}_{t-1} + \boldsymbol{\theta_4} \Delta x_t \mathbf{z}_{t-1} + \boldsymbol{\theta_8} \mathbf{w} + \epsilon_t, \quad (12)$$

where y is average per capita daily caloric consumption, x is GDP per capita, $\mathbf{z}$ is a vector of regime type indicators, $\mathbf{w}$ is a vector of country fixed effects,[18] and ϵ is idiosyncratic error. Since countries rarely change regime types, B & K do not include multiple periods of $\mathbf{z}$. As a result, only one period of regime type is interacted with growth, forcing $\theta_2 = \theta_5 = \theta_6 = \theta_7 = 0$. In addition to this restricted specification, I also estimate the general model in Equation 3 but with the country dummies retained.

Before estimation, I pretest the data to ensure that the ECM is appropriate. I examine the order of integration for each series, as well as the interaction between growth and regime type, using ADF, KPSS, and Ljung-Box tests, and visually inspect Autocorrelation Function (ACF) and partial ACF (PACF) plots. These diagnostics are complicated by the structure of the data (an unbalanced panel with short time series), but the preponderance of evidence indicates that the interaction is a unit root process. Thus, the interaction is integrated, and—by Proposition 2—cointegrated with the other variables, which I confirm using the Engle-Granger and Johansen cointegration tests. As a whole, these results suggest that the ECM and ADL are appropriate models for these data, and OLS estimates will be consistent.[19]

Table 3: Replication of Blaydes and Kayser (2011): comparing the restricted and general models

	B & K model	General model
Calories_{t-1}	-0.08^*	-0.08^*
	(0.01)	(0.01)
ΔGDP_t	3.51^*	0.26
	(1.43)	(3.28)
GDP_{t-1}	0.31	0.49
	(0.30)	(0.43)
$\Delta\text{Regime type}_t$		-6.30
		(3.27)
Hybrid_{t-1}	12.95^*	18.45^*
	(3.88)	(4.84)
Democracy_{t-1}	9.67	11.08
	(5.29)	(7.30)
$\Delta\text{GDP}_t \times \text{Hybrid}_{t-1}$	9.96^*	11.84^*
	(3.15)	(3.16)
$\Delta\text{GDP}_t \times \text{Democracy}_{t-1}$	10.50^*	10.54^*
	(2.50)	(2.51)
$\text{GDP}_{t-1} \times \Delta\text{Regime type}_t$		0.13
		(0.17)
$\Delta\text{GDP}_t \times \Delta\text{Regime type}_t$		2.94
		(2.84)
$\text{GDP}_{t-1} \times \text{Hybrid}_{t-1}$		-0.63^*
		(0.29)
$\text{GDP}_{t-1} \times \text{Democracy}_{t-1}$		-0.25
		(0.32)
Observations	3,264	3,264
R^2	0.10	0.11
RMSE	77.83	77.32

$^*p < .05$. The dependent variable is ΔCalories. Both models include 112 country fixed effects, with heteroskedastic-robust standard errors. RMSE is calculated from out-of-sample predictions using 5-fold cross-validation.

5.2 Results

Results from these models are presented in Table 3. Column 2 replicates the headline results from Blaydes and Kayser (2011) and column 3 gives the estimates for the general model.[20] The results appear generally consistent across models, with only the coefficient on economic growth changing significance. Yet they produce very different predicted effects. B & K report that GDP growth of $\$100$ increases caloric consumption for all countries, with the effect significantly larger in hybrid regimes and democracies. In contrast, Figure 4 plots instantaneous effects for the general model, holding each country's regime type constant; here we cannot distinguish instantaneous effects across regime types, despite the large and significant coefficients on the interaction terms. All that can be said is that there is some evidence that growth is associated with greater caloric consumption among the poor in hybrid regimes and democracies.

More differences between the models' predictions emerge in the dynamic effects. The top and bottom panels of Figure 5 plot period-specific and cumulative effects, respectively. Examining the period-specific effects, we have essentially no information except that, following an initial shock, the effect of growth on caloric consumption is negative in hybrid regimes. However, the trend is very similar in democracies, so it is difficult to draw sharp distinctions among regime types. Turning now to cumulative effects, we see that the instantaneous and long-run effects overlap for each regime. None of the LREs—6 calories in autocracies, -2 in hybrid regimes, and 3 in democracies—are statistically distinguishable from zero. These results contrast sharply with those of B & K, who report LREs of 13, 170, and 187, respectively. Additionally, the LREs overlap across all regimes, meaning that it is not possible to infer whether the effect of growth

[17] B & K omit five observations with outlier values for growth, each more than seven standard deviations from the mean, which I also drop. 69 further observations drop out due to missingness; to facilitate model comparison, I use the remaining sample to estimate all specifications. I also follow the authors in providing heteroskedastic-robust standard errors without clustering.

[18] Adding fixed effects to a model that includes the lagged dependent variable likely induces Nickell (1981) bias. I include them here to maintain comparability across studies, since B & K use them to account for cross-sectional heterogeneity. Note, however, that my results are robust to estimating the general model without country fixed effects.

[19] See the online appendix for full details of all diagnostics.

[20] Estimates for the B & K model reported here differ from the published results because of the slightly smaller sample, but nothing changes in substantive or statistical significance.

on caloric consumption differs across regime types.

Threshold effects provide slightly more leverage. Setting $\lambda = 0$ and visually inspecting the bottom panel of Figure 5, it is clear that the lower bound of the estimated cumulative effect crosses zero eight years after the initial shock in hybrid regimes, but only after 12 years in democracies. Thus, the positive association between growth and calories lasts about 60% longer in democracies than it does in hybrid regimes. This may be suggestive of the clientelism mechanism B & K introduce, since we would expect the effect of non-programmatic redistribution to fade over time. Yet it is difficult to read too much into this result, given the null findings within and across periods and regime types.

Finally, unlike B & K's model, the general model predicts effect sizes that begin and remain small. Mean daily consumption across the sample is 2,405 calories, so $\$100$ in GDP growth generates only a 0.5% increase in caloric intake. For a country to move from the third decile to the cross-national mean in calories consumed, it would require growth of approximately \$1,700 per capita—more than doubling the mean

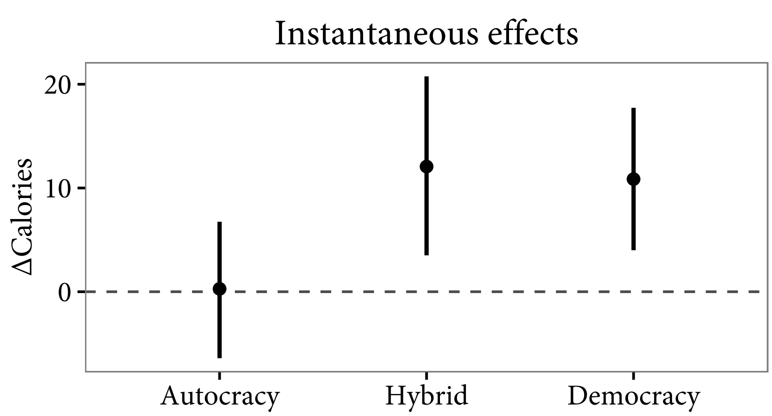

Figure 4: Short-run relationships between growth and inequality by regime type. Each point represents the predicted instantaneous effect on average daily per capita calories consumed of a $\$100$ increase in per capita GDP, conditional on regime type, from the full model. 95% confidence intervals constructed from quantiles of 5,000 samples from $\mathcal{N}_{MV}\left(\hat{\boldsymbol{\theta}}, \mathbb{C}(\hat{\boldsymbol{\theta}})\right)$ are plotted as lines.

income level. No matter the regime type, it appears that growth engenders negligible redistribution.

Taken as a whole, these findings suggest that there are few cross-regime differences in translating growth into inequality reduction. Hybrid regimes and democracies may be marginally better than autocracies at generating immediate caloric gains for the poor, but effect sizes are small, and grow smaller still over time. In the long run, regime types are essentially indistinguishable in their capacity for pro-poor growth.

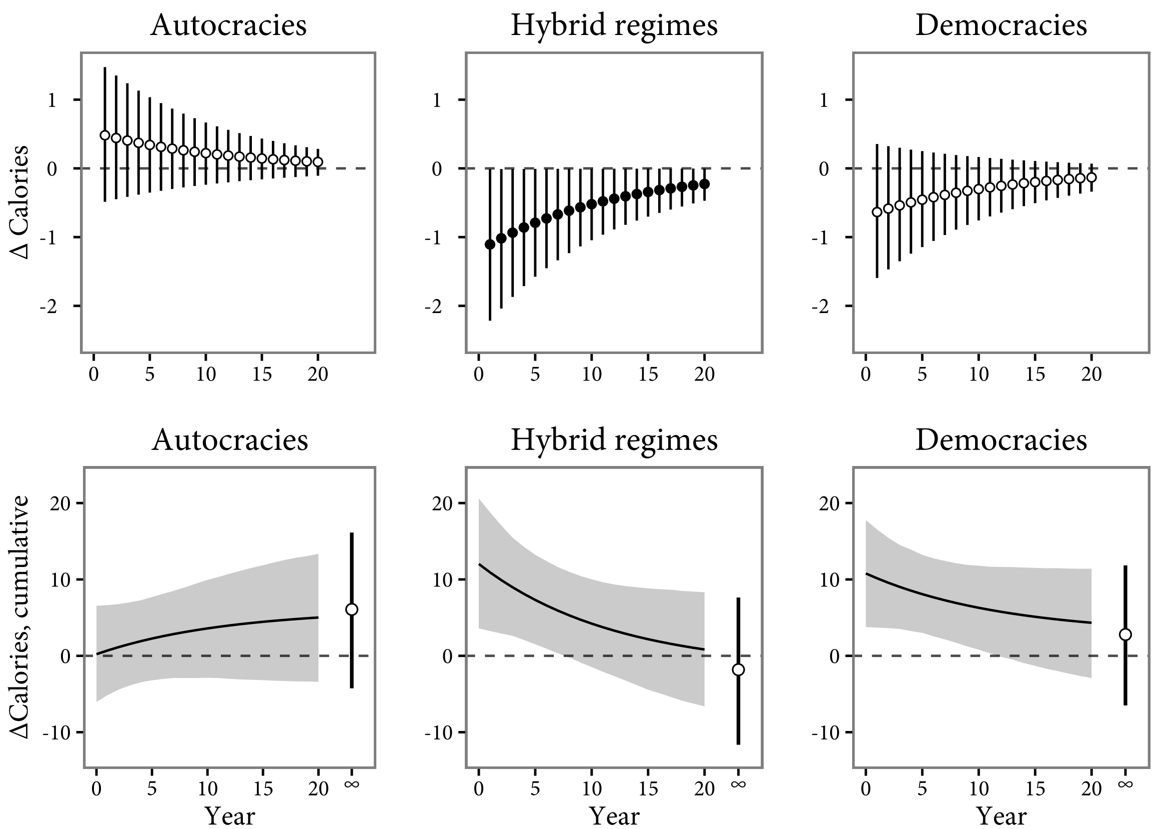

Figure 5: Dynamic relationships between growth and inequality by regime type. The top (bottom) panel plots period-specific (cumulative) effects in average daily per capita calories consumed as the result of \$100 growth in per capita GDP, conditional on regime type, from the full model. LREs (total effects) are plotted at ∞. Lines (ribbons) represent 95% confidence intervals constructed from quantiles of 5,000 samples from $\mathcal{N}_{MV}\left(\hat{\boldsymbol{\theta}}, \mathbb{C}(\hat{\boldsymbol{\theta}})\right)$.

This interpretation contrasts with that of Blaydes and Kayser (2011), who conclude that "growth in autocracies benefits the poor less than growth in democracies and hybrid regimes" (902). These results suggest that theories of regime transition may need to reconsider the assumption that democracies redistribute more than non-democracies.

6 Conclusion

Scholars are increasingly generating theories that rely on conditional relationships. In the effort to "take time seriously" (De Boef and Keele 2008), they have built vast datasets and estimated ADLs and ECMs that capture dynamic political phenomena. However, our understanding of these models has failed to keep pace with the richness of our theories. A number of recent studies have added interactions to ADLs and ECMs, but have not begun the work of developing a theoretical and practical basis for such extensions.

This paper provides first steps in both. I have outlined the conditions necessary for OLS to guarantee consistent estimates, a method for writing general models with interactions, a unified approach for drawing inferences from these models, and simulation evidence that shows the advantages of this approach over current scholarly practice. Revisiting Blaydes and Kayser (2011), I find that the authors' inferences about growth and inequality suffer from two problems: biased estimates produced by models with invalid parameter restrictions, and the application of incorrect formulas for computing predicted effects. A more general framework suggests that democratic institutions do not promote pro-poor growth. These results are troubling for canonical theories in political economy that assume important roles for political institutions in mediating the distributional effects of economic growth.

I conclude by offering a tentative set of "best practices" for scholars testing theories of conditional relationships with dynamic models.

1. *Estimation.* Scholars should ensure that the ADL or ECM produces consistent estimates. Due diligence with respect to diagnosing properties of political time series, already a key step in studying dynamic models, is even more important when complications such as multiplicative interactions are introduced. To this

end, scholars should extend standard pretesting and postestimation procedures to interactions terms.

2. *Specification.* Once assured that their estimates are consistent, scholars should then estimate a general model before exploring restricted models. Theories of how variables interact over time are better tested empirically than imposed *ad hoc*, since invalid parameter restrictions introduce bias and threaten inferences.

3. *Interpretation.* Whatever specification is chosen for interpretation, scholars should use the general approach outlined above to derive quantities of interest *for that model.* Relying on off-the-shelf equations, particularly for the LRE, can lead to incorrect inferences. These well-known formulas are special cases, appropriate only for particular specifications. Scholars should then report those quantities that provide the most information about how the dynamic system behaves. In some cases, this will entail reporting the LRE and mean and median lag lengths. However, in many cases, other quantities such as finite-period cumulative effects and absolute thresholds will be more informative. Whatever estimated effects are discussed, they should be presented with uncertainty statements generated through such techniques as parametric bootstrapping.

These practices will both expand the range of inferences we are able to make and increase our confidence in them, contributing to a stronger evidential base for political science theories.

Appendix

Recall that the general model with a conditional relationship is given in ADL form in Equation 2 and in Equation 3 in ECM form. All of the quantities described in the text can be recovered from either specification; Table A1 provides the variable translations to move between models. These can be used to rewrite the general equations for quantities of interest from the ECM, which I provide here.

Period-specific effects: Equation 4 becomes

$$\frac{\partial y_{t+j}}{\partial x_t} = \begin{cases} 0 & \text{for } j \in \mathbb{Z}_{<0}, \\ \theta_0 + \theta_4 z_t + (\theta_5 + \theta_6)\,\Delta z_t & \text{for } j = 0, \\ (\gamma_1 + 1)^j \left[\theta_0 + \theta_4 z_t + (\theta_5 + \theta_6)\,\Delta z_t\right] + & \\ \qquad (\gamma_1 + 1)^{j-1}\left[\theta_1 - \theta_0 - \theta_4 z_{t+1} - \theta_6 \Delta z_{t+1} + \theta_7 z_t\right] & \text{for } j \in \mathbb{Z}_{>0}. \end{cases} \tag{A1}$$

Cumulative effects: Equation 5 yields

$$\sum_{j=h}^{k} \frac{\partial y_{t+j}}{\partial x_t} = \frac{(\theta_0 + \theta_4 z_t + \theta_5 \Delta z_t + \theta_6 \Delta z_t)\left([\gamma_1 + 1]^h - [\gamma_1 + 1]^{k+1}\right)}{-\gamma_1} + \frac{(\theta_1 - \theta_0 - \theta_4 z_{t+1} - \theta_6 \Delta z_{t+1} + \theta_7 z_t)\left([\gamma_1 + 1]^{h-1} - [\gamma_1 + 1]^{k}\right)}{-\gamma_1}$$

for $h, k \in \mathbb{Z}_{>0}$ such that $h < k$, and

$$\sum_{j=h}^{k} \frac{\partial y_{t+j}}{\partial x_t} = \frac{(\theta_0 + \theta_4 z_t + \theta_5 \Delta z_t + \theta_6 \Delta z_t)\left(1 - [\gamma_1 + 1]^{k+1}\right)}{-\gamma_1} + \frac{(\theta_1 - \theta_0 - \theta_4 z_{t+1} - \theta_6 \Delta z_{t+1} + \theta_7 z_t)\left(1 - [\gamma_1 + 1]^{k}\right)}{-\gamma_1}$$

Table A1: Variable transformations between the general ADL and ECM in Equations 2 and 3

ADL	ECM
$\alpha_0 = \gamma 0$	$\alpha_0 = \gamma_0$
$\alpha_1 = \gamma_1 + 1$	$\alpha_1 - 1 = \gamma_1$
$\beta_0 = \theta_0$	$\beta_0 = \theta_0$
$\beta_1 = \theta_1 - \theta_0$	$\beta_0 + \beta_1 = \theta_1$
$\beta_2 = \theta_2$	$\beta_2 = \theta_2$
$\beta_3 = \theta_3 - \theta_2$	$\beta_2 + \beta_3 = \theta_3$
$\beta_4 = \theta_4 + \theta_5 + \theta_6$	$\beta_4 - \beta_6 = \theta_4$
$\beta_5 = -(\theta_4 + \theta_6)$	$\beta_4 - \beta_5 = \theta_5$
$\beta_6 = -(\theta_5 + \theta_6)$	$-(\beta_4 + \beta_5 + \beta_6) = \theta_6$
$\beta_7 = \theta_6 + \theta_7$	$\beta_4 + \beta_5 + \beta_6 + \beta_7 = \theta_7$

for $h = 0$. Also note that the LRE is given by

$$\sum_{j=0}^{\infty} \frac{\partial y_{t+j}}{\partial x_t} = \frac{\theta_1 + \theta_7 z_t + (\theta_5 + \theta_6)\,\Delta z_t - (\theta_4 + \theta_6)\,\Delta z_{t+1}}{-\gamma_1}. \tag{A2}$$

Threshold effects: First, for the ADL, define

$$\mathbf{a} = \beta_0 + \beta_4 z_t + \beta_6 z_{t-1}$$
$$\mathbf{b} = \beta_1 + \beta_5 z_{t+1} + \beta_7 z_t$$

for ease of presentation. Also define an arbitrary threshold λ, and solve for k from Equation 6 where $h = 0$. This yields

$$k = \left\lceil \frac{\log\left(\dfrac{\mathbf{a} + \mathbf{b} - (1 - \alpha_1)\,\lambda}{\alpha_1 \mathbf{a} + \mathbf{b}}\right)}{\log \alpha_1} \right\rceil,$$

where the ceiling function indicates rounding up since k is a discrete period in time. Lag lengths are identically derived, where $\lambda \equiv \delta\Omega$, $\delta \in [0, 1)$, and Ω is the LRE. As discussed in the text, both quantities are more easily found inductively than solved for.

Equivalently, for the ECM, define

$$\begin{aligned}\mathbf{a} &= \theta_0 + \theta_4 z_t + (\theta_5 + \theta_6)\,\Delta z_t \\ \mathbf{b} &= \theta_1 - \theta_0 - \theta_4 z_{t+1} - \theta_6 \Delta z_{t+1} + \theta_7 z_t.\end{aligned}$$

Threshold effects are then given by

$$k = \left\lceil \frac{\log\left(\dfrac{\mathbf{a} + \mathbf{b} + \gamma_1 \lambda}{(\gamma_1 + 1)\mathbf{a} + \mathbf{b}}\right)}{\log(\gamma_1 + 1)} \right\rceil.$$

References

Acemoğlu, Daron, and James A. Robinson. 2006. *Economic Origins of Dictatorship and Democracy*. Cambridge: Cambridge University Press.

Aklin, Michaël, and Johannes Urpelainen. 2013. "Political Competition, Path Dependence, and the Strategy of Sustainable Energy Transitions." *American Journal of Political Science* 57 (3): 643-658.

Alexiadou, Despina. 2015. "Ideologues, Partisans, and Loyalists: Cabinet Ministers and Social Welfare Reform in Parliamentary Democracies." *Comparative Political Studies* 48 (8): 1051-1086.

Andersen, Jørgen J., and Michael L. Ross. 2014. "The Big Oil Change: A Closer Look at the Haber-Menaldo Analysis." *Comparative Political Studies* 47 (7): 993-1021.

Arnold, R. Douglas, and Nicholas Carnes. 2012. "Holding Mayors Accountable: New York's Executives from Koch to Bloomberg." *American Journal of Political Science* 56 (4): 949-963.

Beck, Nathaniel. 1989. "Estimating Dynamic Models Using Kalman Filtering." *Political Analysis* 1 (1): 121-156.

Blaydes, Lisa, and Mark Andreas Kayser. 2011. "Counting Calories: Democracy and Distribution in the Developing World." *International Studies Quarterly* 55 (4): 887-908.

Boix, Carles. 2003. *Democracy and Redistribution*. Cambridge: Cambridge University Press.

Box, George E. P., and G. Jenkins. 1970. *Time-Series Analysis: Forecasting and Control*. San Francisco: Holden Day.

Box-Steffensmeier, Jan, John R. Freeman, Jon Pevehouse, and Matthew Perry Hitt. 2014. *Time Series Analysis for the Social Sciences*. New York: Cambridge University Press.

Carlin, Ryan E., Gregory J. Love, and Cecilia Martínez-Gallardo. 2015. "Security, Clarity of Responsibility, and Presidential Approval." *Comparative Political Studies* 48 (4): 438-463.

Casillas, Christopher J., Peter K. Enns, and Patrick C. Wohlfarth. 2011. "How Public Opinion Constrains the U.S. Supreme Court." *American Journal of Political Science* 55 (1): 74-88.

Chang, Eric C.C. 2008. "Electoral Incentives and Budgetary Spending: Rethinking the Role of Political Institutions." *Journal of Politics* 70 (4): 1086-1097.

Cheon, Andrew, and Johannes Urpelainen. 2015. "Escaping Oil's Stranglehold: When Do States Invest in Energy Security." *Journal of Conflict Resolution* 59 (6): 953-983.

Davidson, Russell, and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.

De Boef, Suzanna, and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52 (1): 184-200.

Deininger, Klaus, and Lyn Squire. 1996. "A New Data Set Measuring Income Inequality." *The World Bank Economic Review* 10 (3): 565-591.

Engle, Robert F., and C. W. J. Granger. 1987. "Co-integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55 (2): 251-276.

Enns, Peter K., Nathan J. Kelly, Jana Morgan, Thomas Volscho, and Christopher Witko. 2014. "Conditional Status Quo Bias and Top Income Shares: How U.S. Political Institutions Have Benefited the Rich." *Journal of Politics* 76 (2): 289-303.

Enns, Peter K., Nathan J. Kelly, Takaaki Masaki, and Patrick C. Wohlfarth. 2016. "Don't Jettison the General Error Correction Model Just Yet: A Practical Guide to Avoiding Spurious Regression with the GECM." *Research and Politics* 3 (2).

Enns, Peter K., Takaaki Masaki, and Nathan J. Kelly. 2014. "Time Series Analysis and Spurious Regression: An Error Correction." Paper presented at the Annual Meeting of the Southern Political Science Association, Orlando, FL. Available at http://takaakimasaki.com/wp-content/uploads/2014/08/EnnsMasakiKelly_ECM_9.25.14.pdf.

Esarey, Justin. 2016. "Fractionally Integrated Data and the Autodistributed Lag Model: Results from a Simulation Study." *Political Analysis* 24 (1): 42-49.

Escribà-Folch, Abel. 2012. "Authoritarian Responses to Foreign Pressure: Spending, Repression, and Sanctions." *Comparative Political Studies* 45 (6): 683-713.

Freeman, John R. 2016. "Progress in the Study of Nonstationary Political Time Series: A Comment." *Political Analysis* 24 (1): 50-58.

Gandrud, Christopher, Laron K. Williams, and Guy D. Whitten. 2016. "dynsim: Dynamic Simulations of Autoregressive Relationships." R package version 1.2.2.

Granger, C. W. J. 1986. "Developments in the Study of Cointegrated Economic Variables." *Oxford Bulletin of Economics and Statistics* 48 (3): 213-228.

Grant, Taylor, and Matthew J. Lebo. 2016. "Error Correction Methods with Political Time Series." *Political Analysis* 24 (1): 3-30.

Helgason, Agnar Freyr. 2016. "Fractional Integration Methods and Short Time Series: Evidence from a Simulation Study." *Political Analysis* 24 (1): 59-68.

Hobolt, Sara Binzer, and Robert Klemmensen. 2008. "Government Responsiveness and Political Competition in Comparative Perspective." *Comparative Political Studies* 41 (3): 309-337.

Houle, Christian. 2009. "Inequality and Democracy: Why Inequality Harms Consolidation but Does Not Affect Democratization." *World Politics* 61 (4): 589–622.

Jennings, Will, and Peter John. 2009. "The Dynamics of Political Attention: Public Opinion and the Queen's Speech in the United Kingdom." *American Journal of Political Science* 53 (4): 838-854.

Jensen, Carsten, and Peter B. Mortensen. 2014. "Government Responses to Fiscal Austerity: The Effect of Institutional Fragmentation and Partisanship." *Comparative Political Studies* 47 (2): 143-170.

Johansen, Soren. 1988. "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12 (2-3): 231-254.

Kayser, Mark Andreas. 2009. "Partisan Waves: International Business Cycles and Electoral Choice." *American Journal of Political Science* 53 (4): 950-970.

Keele, Luke, Suzanna Linn, and Clayton McLaughlin Webb. 2016. "Erratum for Keele, Linn, and Webb (2016)." *Political Analysis* 24 (2): 291-306.

Kellam, Marisa, and Elizabeth A. Stein. 2015. "Silencing Critics: Why and How Presidents Restrict Media Freedom in Democracies." *Comparative Political Studies* 49 (1): 36-77.

Keller, Eric, and Nathan J. Kelly. 2015. "Partisan Politics, Financial Deregulation, and the New Gilded Age." *Political Research Quarterly* 68 (3): 428-442.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44 (2): 341-355.

Kono, Daniel Y. 2008. "Democracy and Trade Discrimination." *Journal of Politics* 70 (4): 942-955.

Kono, Daniel Yuichi, and Gabriella R. Montinola. 2013. "The Uses and Abuses of Foreign Aid: Development Aid and Military Spending." *Political Research Quarterly* 66 (3): 615-629.

Kono, Daniel Yuichi, and Gabriella R. Montinola. 2015. "Foreign Aid, Time Horizons, and Trade Policy." *Political Research Quarterly* 48 (6): 788-819.

Kono, Daniel Yuichi, Gabriella R. Montinola, and Nicholas Verbon. 2015. "Helping Hand or Heavy Hand? Foreign Aid, Regime Type and Domestic Unrest." *International Political Science Review* 36 (4): 409-424.

Lebo, Matthew J., and Janet M. Box-Steffensmeier. 2008. "Dynamic Conditional Correlations in Political Science." *American Journal of Political Science* 52 (3): 688-704.

Lipsmeyer, Christine S., and Ling Zhu. 2011. "Immigration, Globalization, and Unemployment Benefits in Developed EU States." *American Journal of Political Science* 55 (3): 647-664.

Marshall, Monty G., Keith Jaggers, and Ted Robert Gurr. 2011. "Polity IV Project." Center for Systemic Peace.

Miller, Michael K. 2015. "Elections, Information, and Policy Responsiveness in Autocratic Regimes." *Comparative Political Studies* 48 (6): 691-727.

Morgan, Jana, and Nathan J. Kelly. 2013. "Market Inequality and Redistribution in Latin America and the Caribbean." *Journal of Politics* 75 (3): 672-685.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49 (6): 1417-1426.

Nooruddin, Irfan, and Joel W. Simmons. 2006. "The Politics of Hard Choices: IMF Programs and Government Spending." *International Organization* 60 (4): 1001-1033.

Petris, Giovanni, Sonia Petrone, and Patrizia Campagnoli. 2009. *Dynamic Linear Models with R.* New York: Springer.

Ramirez, Mark D. 2013. "The Policy Origins of Congressional Approval." *Journal of Politics* 75 (1): 198-209.

Sorens, Jason. 2011. "The Institutions of Fiscal Federalism." *Publius: The Journal of Federalism* 41 (2): 207-231.

Swank, Duane. 2006. "Tax Policy in an Era of Internationalization: Explaining the Spread of Neoliberalism." *International Organization* 60 (4): 847-882.

Tenorio, Barbara Zarate. 2014. "Social Spending Responses to Organized Labor and Mass Protests in Latin America, 1970-2007." *Comparative Political Studies* 47 (14): 1945-1972.

Werner, Timothy, and John J. Coleman. 2015. "*Citizens United*, Independent Expenditures, and Agency Costs: Reexamining the Political Economy of State Antitakeover Statutes." *Journal of Law, Economics, & Organization* 31 (1): 127-159.

Wright, Joseph. 2015. "Aid Effectiveness and the Politics of Personalism." *Comparative Political Studies* 43 (6): 735-762.

Yule, G. Udny. 1926. "Why Do We Sometimes Get Nonsense-Correlations between Time-Series? A Study in Sampling and the Nature of Time-Series." *Journal of the Royal Statistical Society* 89 (1): 1-63.