Conditional relationships in dynamic models*

Zach Warner†

March 10, 2019

Abstract

Many political science theories include variables that interact over time. In recent years, scholars have turned to autoregressive distributed lag and error correction models to test these theories. However, adding multiplicative interactions to these workhorse models creates difficulties with (1) knowing when OLS estimates are consistent, (2) writing models without implicit 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 on the relationship among democracy, growth, and inequality.

*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 Chris Arnold, Doug Atkinson, Rikhil Bhavnani, Kevin Fahey, Brian Gaines, Jude Hays, René Lindstädt, Noam Lupu, Melanie Manion, Jon Pevehouse, and Alex Tahk. Previous versions of this paper were presented at Cardiff, Wisconsin, and the annual meeting of the Society for Political Methodology.

†Postdoctoral Research Fellow, Centre for Political and Legal Analytics, Cardiff University. email: WarnerZ@cardiff.ac.uk, web: http://www.zachwarner.net.

Political scientists are increasingly developing theories in which 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. Over the past decade, scholars 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 even for workhorse specifications such as autoregressive distributed lag and error correction models (ADLs and ECMs, respectively), there is little guidance on how best to model these interactions. As a result, scholars' findings are vulnerable to three 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 (e.g., Philips 2018). However, it is unclear what tests an interaction term must satisfy, under what circumstances, and so scholars may not be able to guarantee that ordinary least-squares (OLS) will produce consistent estimates.

1 . Other examples include Adams, Ezrow, and Wlezien (2016), Aklin and Urpelainen (2013), Alexiadou (2015), Andersen and Ross (2014), Arnold and Carnes (2012), Blaydes and Kayser (2011), Carlin, Love, and Martínez-Gallardo (2015), Caughey and Warshaw (2018), Chang (2008), Cheon and Urpelainen (2015), Copelovitch, Gandrud, and Hallerberg (2018), Cronert (2018), Dorsch and Maarek (forthcoming), Enns et al. (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, Montinola, and Verbon (2015), Lipsmeyer and Zhu (2011), Meyer and Biegert (2019), Miller (2015), Morgan and Kelly (2013), Sorens (2011), Ramirez (2013), Swank (2006), Tenorio (2014), Valdini and Lewis-Beck (2018), Werner and Coleman (2015), and Wright (2015). See also dynsim, a package for R and Stata that provides a user-friendly resource for studying dynamic models with interactions (Gandrud, Williams, and Whitten 2016).

2. *Specification.* Typically, each variable of interest in an ADL or ECM suffers 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 (Box-Steffensmeier et al. 2014). 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 together allow scholars to interpret their estimates. However, interactions complicate the computation of these quantities, including both short- and long-run effects. Scholars may be unaware of how to derive quantities of interest for their models, potentially leading 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] 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.

I then demonstrate 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.

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

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

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 turn to Monte Carlo simulation 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, Linn, and Webb 2016). Second, we do not know whether these risks outweigh the potential for bias that arises from invalid restrictions. Despite these risks, I find that the approach introduced here produces more reliable inferences: while there is some evidence of overfitting, this risk is relatively minimal compared to the cost of an invalid restriction on the model. 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.[3] 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 approach developed here. I find that the authors' conclusions suffer from biased estimates: 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.

3 . The Appendix also includes replications of Morgan and Kelly (2013) and Jennings and John (2009), with similar results.

1 Ensuring consistent estimates

There are a great many ways of modeling dynamic dependence among variables. Due to space constraints, 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).[4] Throughout the paper, I develop my argument with the ADL because it facilitates transparency, but my results also hold for the ECM because the two are simple linear reparameterizations of one another (Davidson and MacKinnon 1993; De Boef and Keele 2008).[5]

The ADL can be written in its 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.[6] Intuitively, we want to

4 . 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, Petrone, and Campagnoli (2009).

5 . In the Appendix, I provide derivations for the inferential framework I introduce in Section 3, as well as derivations for this same framework represented as an ECM. 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.

6 . Throughout this paper I refer only to covariance (or weak) stationarity. See the 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.

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.[7]

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.[8]

Ensuring that a model with a conditional relationship satisfies these requirements is not trivial. It seems intuitive that the "memory" of xz would directly follow the properties of x and z. Yet this relationship does not always hold, and the standard practice of diagnosing the time-series properties of all lower-order terms cannot guarantee consistent OLS estimates. On the other hand, 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.[9] 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.

7 . 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.

8 . All ADLs and ECMs must satisfy an additional requirement: equation balance. Recent studies have debated how best to ensure equation balance (Enns et al. 2016; Grant and Lebo 2016; Keele, Linn, and Webb 2016). Due to space constraints, I leave the role of equation balance in ADLs and ECMs with interactions for future research.

9 . Proofs are provided in the Appendix.

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 (Yule 1926; Grant and Lebo 2016).

On one hand, the result in Proposition 1 adds to the already-substantial burden on researchers to pretest 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 (for a recent review, see Box-Steffensmeier et al. 2014).[10] 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.

Results are 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. 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 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. Further, there is much less information in time series data than sample sizes suggest, making it likely that scholars who add extra covariates are simply

10 . Grant and Lebo (2016) demonstrate the utility of fractional integration (FI) methods for data that may be neither stationary nor integrated. I leave the question of interactions in FI models for future research.

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

modeling noise. Results from such specifications will be unstable (Keele, Linn, and Webb 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 simulation analysis below.

2 Specifying 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 often 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 may generate incorrect or incomplete interpretations.

The simplest way to prevent these issues from arising is to estimate a general model that allows a covariate interaction to unfold freely across time. The most general ADL that captures this relationship between two variables, each autoregressive of order 1, or AR(1), is given by

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

This model 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.[12] 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. For one, 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 it is non-zero in the true DGP biases estimates of all xz terms, and through the conditional relationship, all x and z terms (that is, it biases *all* β terms in Equation 2). The Monte Carlo exercise below underscores the magnitude of these problems: in some cases, invalid restrictions can produce essentially random estimates.

This general model is also 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.

A final advantage of this model is it 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.

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

3 Interpreting dynamic models

Scholars studying dynamic models are typically interested in change in the response y as a function of change in (or "shock" to) a variable x. These quantities of interest fall into three categories: period-specific effects, cumulative effects, and threshold effects.[13]

The first of these quantities are period-specific effects, 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

13 . 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.

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, Enns, and Wohlfarth (2011) examine the effect of public opinion on Supreme Court decisions. Computing period-specific effects from their estimates reveals that 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. Rather than monotonically increasing over time, the effect of public opinion on judicial decisions is immediately large but diminishes significantly. Studying period-specific effects can uncover nuanced relationships, improving our understanding of dynamic systems.

The second quantity of interest, often the motivation for studying such models, are cumulative effects. These effects, which provide information about relationships over time, 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. Just as 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, so too is $t = 0$ a special case for 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:

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

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.

Finite-period cumulative effects are useful for answering many questions scholars care about which the LRE cannot. Most notable are questions about fixed time periods such as a presidential term. Using the LRE to describe such relationships is unreliable because the difference between the cumulative effect from zero to k and from zero to infinity is often non-trivial.[14] 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. Even very large

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

infinite-period relationships are not necessarily good approximations of finite-period relationships.

Further, finite-period cumulative effects are arguably more interesting and substantively important than the LRE. For instance, the question "What is the effect of an economic shock on public opinion over the length of a presidential term?" cannot be meaningfully 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, particularly where data constraints bind. Examining cumulative effects over various time windows can ameliorate the problem of extrapolating from short time series, and help scholars make more precise statements about the substantive significance of their findings.

The last quantity of interest that can be recovered from Equation 2 are threshold effects, 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.[15] However, 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.

Whether they are interested in period-specific, cumulative, and threshold effects, scholars may wish to compare results from the general model with those from a "restricted" specification, in which one or more of the $(p+1)^2$ interaction terms are excluded. While such comparisons are useful in many contexts, scholars may not be aware of the drastic consequences these parameter restrictions pose for calculating the quantities discussed above (by implicitly setting β or θ terms to zero). In general, quantities derived from one model do not match those from another model, even if only a single term differs between them.

Kono and Montinola (2013) provide an example. The authors 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

15 . 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, Enns, and Wohlfarth (2011) example above. The inferential value of this quantity is therefore limited.

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.[16] Setting aside regime change could therefore lead to incomplete or incorrect conclusions about the dynamic system. The only way to protect inferences from such threats is to derive quantities of interest *for each specific model*.

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.[17] This approach offers significant advantages over the standard practice for estimating uncertainty, the "delta method," which is not easily extended to complex models nor to all of these quantities of interest. In practice, this often leads to interpreting point estimates in the absence of uncertainty bounds,[18] or only interpreting instantaneous effects and the LRE. Parametric bootstrapping resolves this problem. 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.

4 Simulation evidence

Although the general model introduced above has a number of attractive features, it does raise the question of whether these $(p+1)^2$ interaction parameters induce ECR bias and overfitting. If the benefits of fitting a general model are outweighed by these problems, then scholars may wish to study less complex models with implicit parameter restrictions instead. I use Monte Carlo simulation to answer this question.

16 . See the Appendix for the LRE of an unrestricted ECM.
17 . See the Appendix for a more detailed discussion of this procedure.
18 . See the interpretation of lag lengths in De Boef and Keele (2008).

I first 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 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,

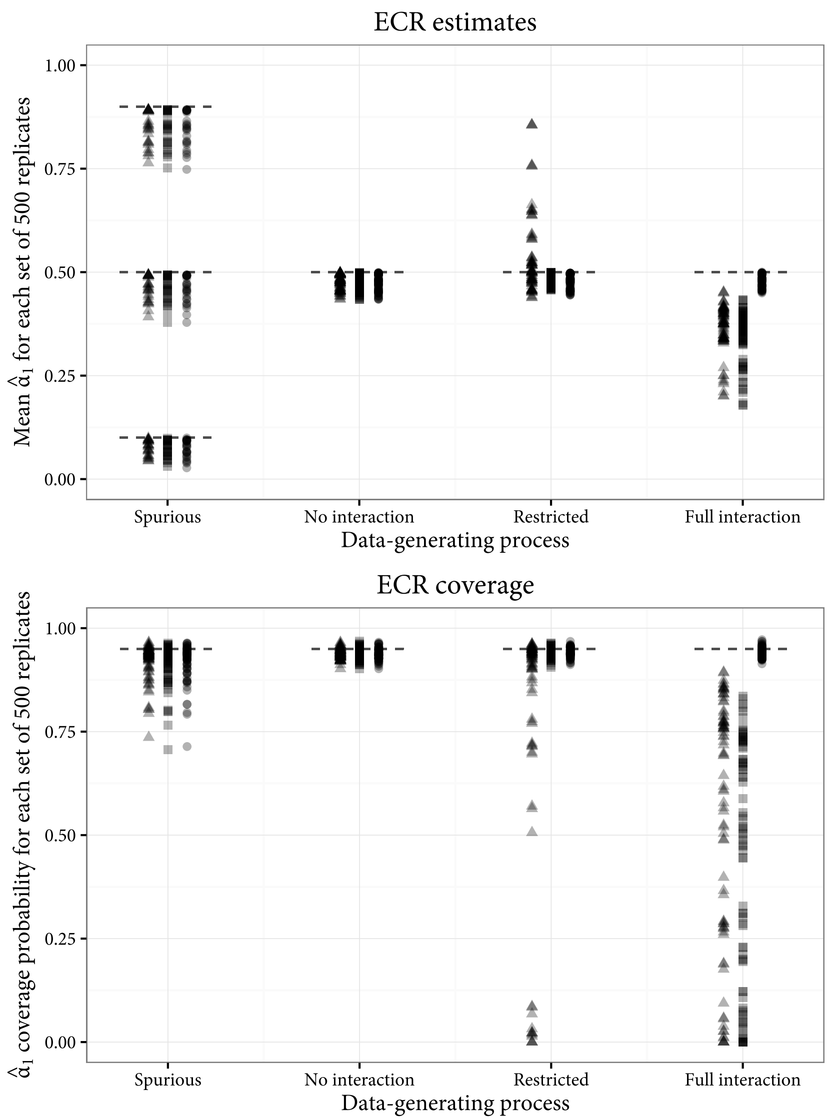

Figure 1: Monte Carlo simulation results for the error-correction rate, α_1. Triangles represent estimates from the model without an interaction, squares are from a restricted model, and circles are from the general model. Mean estimates are plotted against the true values (dashed lines) in the top panel. 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

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.

Turning now to overfitting: if it 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.[19] 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 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

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

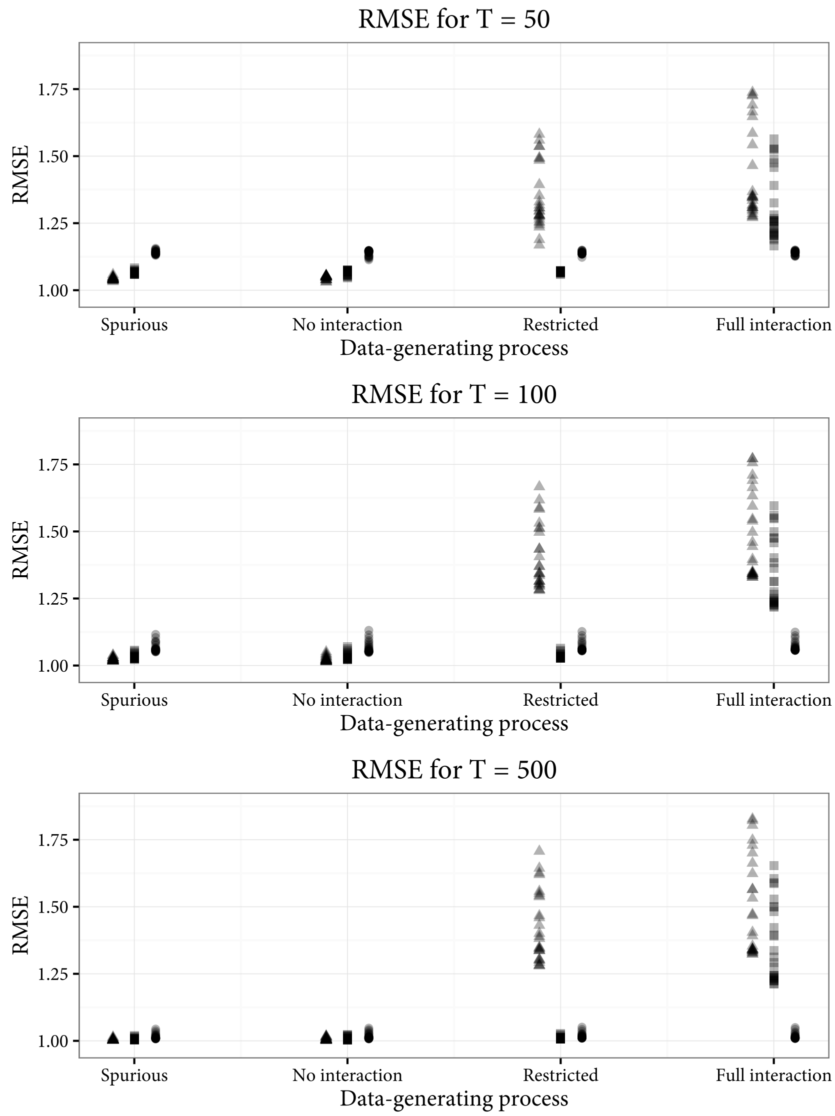

Figure 2: Monte Carlo simulation results for RMSE from 5-fold cross-validation. Triangles represent RMSE from the model without an interaction, squares are from a restricted model, and circles are from the general model. The top panel is for a time series of length $T = 50$, the middle is $T = 100$, 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

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.

Last is the question of tradeoffs: 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.

Taken together, these results indicate that 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, as political scientists almost always study restricted models without comparing them against a more general specification.[20] My findings indicate that this current standard operating procedure likely produces seriously biased estimates and unreliable inferences.

20 . The only exceptions of which I am aware are Copelovitch, Gandrud, and Hallerberg (2018), Cronert (2018), and Meyer and Biegert (2019).

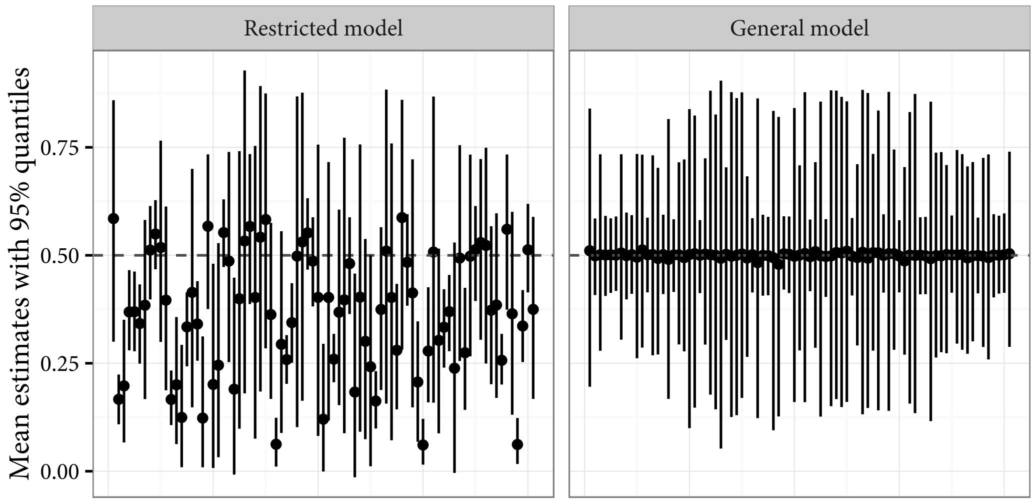

Figure 3: The consequences of invalid restrictions. Monte Carlo simulation results for β_4 when the DGP includes a complex conditional relationship. Points represent mean estimates from 500 replicates of each experimental condition, with lines for 95% quantiles.

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.[21] 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).

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

21 . 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 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.

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.

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, Jaggers, and Gurr 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.[22]

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

$$\Delta y_t = \gamma_0 + \gamma_1 y_{t-1} + \theta_0 \Delta x_t + \theta_1 x_{t-1} + \boldsymbol{\theta}_3 \mathbf{z}_{t-1} + \boldsymbol{\theta}_4 \Delta x_t \mathbf{z}_{t-1} + \boldsymbol{\theta}_8 \mathbf{w} + \epsilon_t, \quad (12)$$

22 . B & K omit five observations with outlier values for growth, 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.

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,[23] 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.[24]

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.[25] 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

23 . 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.

24 . See the Appendix for full details of all diagnostics.

25 . 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.

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.

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 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

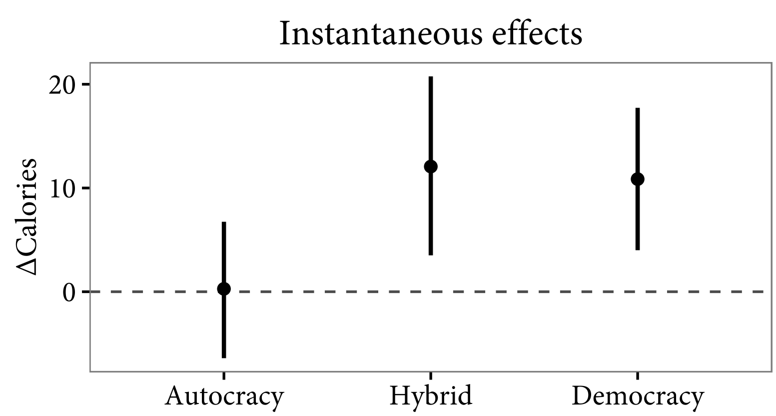

Figure 4: Short-run relationships between growth and inequality by regime type. Each point represents the predicted instantaneous effect on average daily per capita calories consumed of a $\$100$ increase in per capita GDP, conditional on regime type, from the full model. 95% confidence intervals constructed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

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

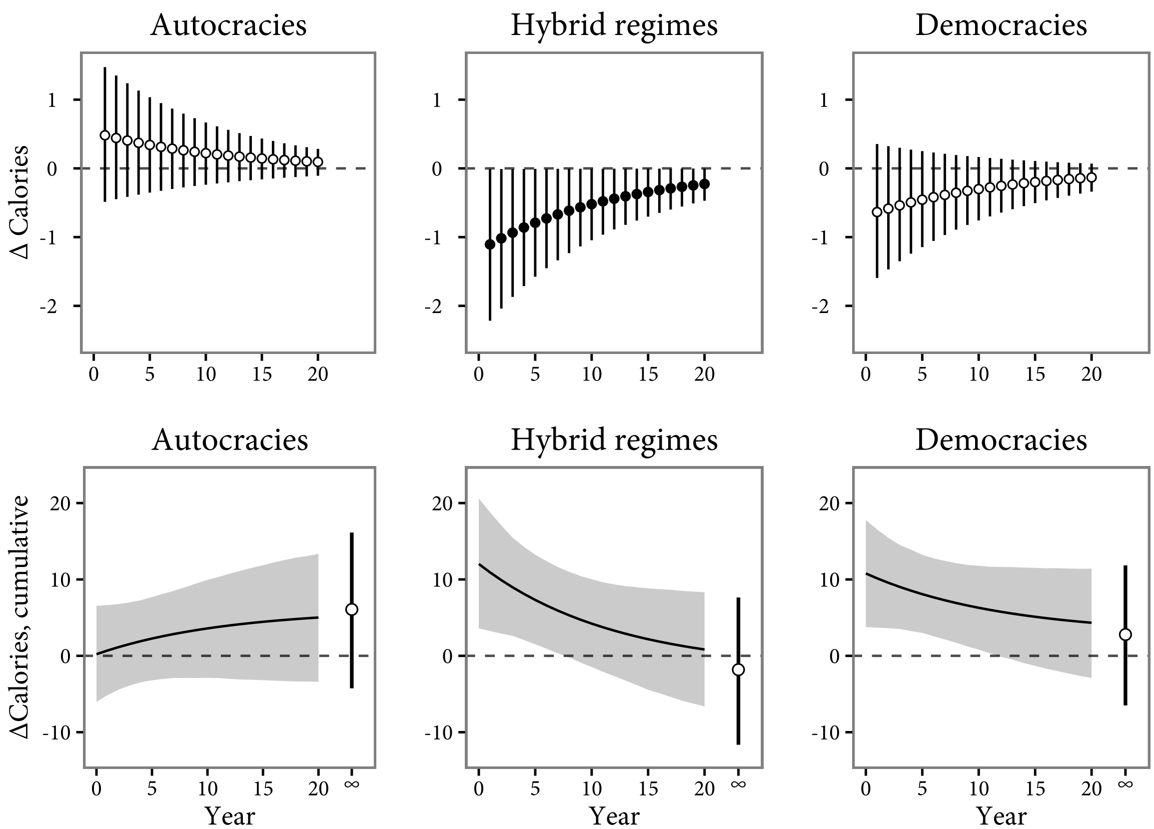

Figure 5: Dynamic relationships between growth and inequality by regime type. The top (bottom) panel plots period-specific (cumulative) effects in average daily per capita calories consumed as the result of $\$100$ growth in per capita GDP, conditional on regime type, from the full model. LREs (total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

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 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. 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 Understanding dynamic interactions

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 biased estimates produced by models with invalid parameter restrictions. 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.

My findings demonstrate that scholars should not unthinkingly add interaction terms to their dynamic models. Here I echo the chorus of findings that applied work typically suffers from too little attention to data constraints, diagnostic tests, and specification decisions.[26] Like Grant and Lebo (2016) and Keele, Linn, and Webb (2016), I urge scholars to rigorously test the properties of their data at each stage of modeling and estimation.

The central contribution of this paper is to provide a toolkit for scholars to extend dynamic models to include conditional relationships. I therefore conclude by offering a tentative set of "best practices" for studying these 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 study a general model before exploring restricted models. Theories of how variables interact over time are better tested empirically than imposed *a priori*, 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

26 . 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), and Keele, Linn, and Webb (2016).

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 understanding how political forces interact over time.

References

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

Adams, James, Lawrence Ezrow, and Christopher Wlezien. 2016. "The Company You Keep: How Voters Infer Party Positions on European Integration from Governing Coalition Arrangements." *American Journal of Political Science* 60 (4): 811–823.

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

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

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.

Caughey, Devin, and Christopher Warshaw. 2018. "Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 1936-2014." *American Political Science Review* 112 (2): 249–266.

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.

Copelovitch, Mark, Christopher Gandrud, and Mark Hallerberg. 2018. "Financial Data Transparency, International Institutions, and Sovereign Borrowing Costs." *International Studies Quarterly* 62 (1): 23–41.

Cronert, Axel. 2018. "Accommodation or Extraction? Employers, the State, and the Joint Production of Active Labor Market Policy." *Politics & Society* 46 (4): 539–569.

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.

Dorsch, Michael T., and Paul Maarek. Forthcoming. "Democratization and the Conditional Dynamics of Income Distribution." *American Political Science Review.*

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, 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., 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., 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.

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.

———. 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.

Meyer, Brett, and Thomas Biegert. 2019. "The Conditional Effect of Technological Change on Collective Bargaining Coverage." *Research and Politics.*

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.

Philips, Andrew Q. 2018. "Have Your Cake and Eat It Too? Cointegration and Dynamic Inference from Autoregressive Distributed Lag Models." *American Journal of Political Science* 62 (1): 230–244.

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.

Valdini, Melody E., and Michael S. Lewis-Beck. 2018. "Economic Voting in Latin America: Rules and Responsibility." *American Journal of Political Science* 62 (2): 410–423.

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.