But Wait, There's More! Maximizing Substantive Inferences from TSCS Models

Laron K. Williams Department of Political Science University of Missouri
Guy D. Whitten Department of Political Science Texas A&M University

Political scientists rarely take full advantage of the substantive inferences that they can draw from time-series cross-section data. Most studies have emphasized statistical significance and other standard inferences that can be drawn from single coefficients over one time period. We show that by simulating the quantities of interest over longer periods of time and across theoretically interesting scenarios, we can draw much richer inferences. In this article, we present a technique that produces graphs of dynamic simulations of relationships over time. Graphical simulations are useful because they represent long-term relationships between key variables and allow for examination of the impact of exogenous and/or endogenous shocks. We demonstrate the technique's utility by graphically representing key relationships from two different works. We also present a preliminary version of the dynsim command, which we have designed to extend the Clarify commands in order to produce dynamic simulations.

In recent years, Time-Series Cross-Sectional (TSCS) data have become more available to political scientists, presenting researchers with an interesting combination of promise and problems. On the one hand, TSCS data allow researchers opportunities to gain variance on theoretically critical variables that do not vary much within individual units over time. But, on the other hand, models estimated from these data can be plagued by a particularly daunting list of statistical obstacles that can lead to faulty hypothesis-testing inferences. Given the challenges posed by the simultaneous occurrence of problems associated with time-series data (e.g., nonstationarity and autocorrelation), problems associated with cross-sectional data (e.g., heteroskedasticity), and problems unique to TSCS data (e.g., contemporaneous error correlation), it is not surprising that political methodologists have spent a lot of time worrying about testing assumptions and estimating believable covariance matrices (Beck and Katz 1995; Green, Kim, and Yoon 2001; Stimson 1985).

We agree with this emphasis on getting believable estimates and avoiding spurious conclusions. But we think that it is important that once an applied researcher has successfully navigated the minefield of statistical obstacles presented by TSCS data, they fully explore the implications of their results. Outside of the TSCS world, political scientists have recognized the importance of simulating the effects of shocks on long-term estimates with the use of methods such as error-correction models (see De Boef and Keele 2008 for a review), vector-autoregression, and graphical interpretations of these shocks with impulse-response functions (see Freeman, Williams, and Lin 1989 for a review). This call to make the most out of complicated estimates is in tune with another set of recent works in political methodology that have called for less reliance on numerical interpretations of statistical significance and more emphasis on simulations to produce helpful graphical depictions of statistical and substantive significance (e.g., King, Tomz, and Wittenberg 2000).

In the remaining sections we begin with a brief overview of common practices for interpreting models estimated with TSCS data. We then make a general case for conducting dynamic simulations by illustrating their usefulness on two political science research applications with the use of an original Stata command called dynsim. We conclude with a discussion of the usefulness of long-term dynamic simulations.

Current Practices

The autoregressive nature of a great deal of political phenomena requires that scholars include a lagged dependent variable for theoretical and/or methodological

The Journal of Politics, Vol. 74, No. 3, July 2012, Pp. 685–693 doi:10.1017/S0022381612000473

ISSN 0022-3816

reasons. As De Boef and Keele (2008) point out, doing so allows scholars to explore a number of quantities of interest that describe dynamic relationships (e.g., long-term effects, long-run equilibrium).[1] A simple review of articles published in the *American Journal of Political Science (AJPS)* and the *American Political Science Review (APSR)* reveals that very few authors produce graphical depictions of these dynamic relationships (12.2% and 15.8%, respectively) and even fewer provide appropriate measures of uncertainty (2% and 10.5%, respectively). We argue that these percentages are too low and that they represent missed opportunities for important substantive inferences.

As an example of these missed opportunities, consider an article published by David Rueda in the *APSR* (Rueda 2005). Rueda advances a compelling argument that, in terms of employment policy preferences, laborers should be divided into those whose employment is relatively secure (insiders) and those who are more vulnerable (outsiders). Because insiders will tend to be the core supporters of Social Democratic parties, Rueda theorizes that we should see increases in employment protection measures during periods of left government. To test this empirical claim, Rueda estimated a TSCS model of severance pay as a function of cabinet partisanship and a series of control variables including a one-year lag of the dependent variable. The coefficient on his main variable of interest, cabinet partisanship, is found to be in the expected direction and statistically significant with a one-tailed p-value of .037. From Table 1, we can see that most authors of the studies that we surveyed (published between 2002 and 2006) stopped at this point in the interpretation of their TSCS models. Rueda, however, correctly interpreted this effect, in the presence of a lagged dependent variable, as the short-term effect of a one-unit change in the independent variable on his dependent variable. He then proceeded to report the long-term effect of this change in the independent variable, estimated from the following formula: $LTE_{X_1} = \frac{\hat{\beta}}{1-\hat{\phi}}$where $\hat{\beta}$ is the parameter estimate for the independent variable of interest and $\hat{\phi}$ is the parameter estimate for the lagged dependent variable. Although this is also a correct interpretation, it is a point estimate for which Rueda did not report any measure of uncertainty.[2]

A particularly helpful way of representing these dynamic relationships is through graphical depictions of the long-run expected values for substantively interesting scenarios. Our version of this type of inferential approach, which we label "dynamic simulation," can be carried out as within-sample inferences (assuming that the values of all of the independent variables are known) or as out-of-sample forecasts (in which we can choose to incorporate a variety of different types of uncertainty).[3] Across the range of choices in terms of

[1]As De Boef and Keele (2008) point out, it is the specification of a lagged dependent variable on the right-hand side of a model that makes it dynamic. Although De Boef and Keele produce a number of insightful graphs for interpretation of dynamic relationships, they do not produce long-run depictions of the dependent variable under different scenarios as we do in with our dynamic simulations. It is also worth noting that while our focus in this article is on TSCS models, all of our discussion applies equally well to models of a single time series. Our Stata command produces simulations for time-series models as well as TSCS models.

[2]One can calculate measures of uncertainty for long-term effects in a number of ways: by simulation techniques or by calculating the analytical standard errors with the following formula: $var\left(\frac{a}{b}\right) = \left(\frac{1}{b^2}\right)var(a) + \left(\frac{a^2}{b^4}\right)var(b) - 2\left(\frac{a}{b^3}\right)cov(a,b)$ where, in this case, $b = 1 - \hat{\phi}$ and $\hat{\phi}$ is the parameter estimate on the lagged dependent variable (De Boef and Keele 2008; 192). Another way to measure the long-term coefficient is by using the Bewley (1979) transformation of the autoregressive distributed lag (ADL) model, which is described in De Boef and Keele (2008; 192). Note that throughout this article and the supplementary materials document, when we refer to the work of other authors, we use their original notation.

[3]Greene (2003, . 571–80) provides an excellent discussion forecasting in the context of autoregressive distributed lag (ARDL) models. He starts with a general formula for ARDL models as follows: $y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{j=0}^{r} \beta_j x_{t-j} + \delta w_t + \varepsilon_t$ where μ is the mean of y, x is an independent variable for which we wish to model the dynamic properties as they pertain to y, w_t is another independent variable whose dynamic properties are appropriately captured by δ, and ε_t is a well-behaved error process that meets the assumptions of OLS. For ease of exposition, Greene collapses the nonendogenous RHS variables to $\mu_t = \mu + \sum_{j=0}^{r} \beta_j x_{t-j} + \delta w_t$ so that the model becomes $y_t = \mu_t + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t$ which simplifies to $y_t = \mu + \gamma_1 y_{t-1} + \ldots + \gamma_p y_{t-p} + \varepsilon_t$. He then provides a nice discussion of the three different sources of error in forecasting, "To form a prediction interval, we will be interested in the variance of the forecast error: $e_{T+1|T} = \hat{y}_{T+1|T} - y_{T+1}$ This error will arise from three sources. First, in forecasting μ_t, there will be two sources of error. The parameters, μ, δ, and $\beta_0, \ldots, \beta_r$ will have been estimated so $\hat{\mu}_{T+1|T}$ will differ from $\hat{\mu}_{T+1}$ because of the sampling variation in these estimators. Second, if the exogenous variables, x_{T+1} and w_{T+1} have been forecasted, then to the extent that these forecasts are themselves imperfect, yet another source of error to the forecast will result. Finally, although we will forecast ε_{T+1} with its expectation of zero, we would not assume that the actual realization will be zero, so this step will be a third sources of error. In principle, an estimate of the forecast error variance $Var\left[e_{T+1|T}\right]$ would account for all three sources of error" (576). However, as Greene goes on to point out, in practice this is nearly impossible. As we noted, we have written our Stata command such that there are options for including forecasting uncertainty such as that discussed by Greene and also the formula provided by Enders (2004). However, since most inferences from statistical models in political science fall into the "within-sample inferences" category, we will continue in this tradition with the dynamic simulations presented here.

Table 1 Percentage of Articles in Leading Journals with Models of TSCS Data that Contain Dynamic Interpretations, 2002–2006

	Journal	
Type of Interpretation	**AJPS**	**APSR**
Standard significance interpretation for individual parameter estimates	100.0%	100.0%
Any type of dynamic interpretation	14.3%	31.6%
Any graphical depiction of dynamic inferences	12.2%	15.8%
Any type of dynamic interpretation with any measure of uncertainty	2.0%	10.5%

error structures (e.g., panel-corrected standard errors vs. OLS), and sources of uncertainty, analytical and simulation-based methods lead to the same dynamic inferences.[4] Our preferred method is to use the Clarify program to simulate expected values over the long-term. While it is not computationally difficult to calculate long-term expected values, it is somewhat tedious to produce the computer code necessary to generate figures that display these calculations over multiple time periods. In order to make these types of long-term dynamic simulations more accessible to scholars, we created a Stata command called dynsim that automates this process. Each predicted value is generated according to the formula $\tilde{y} = X_C\tilde{\beta} + \tilde{\varepsilon}$, where $\tilde{\beta}$ is a vector of simulated effect coefficients, $\boldsymbol{X}_C$ is a matrix of user-specified values of variables, and $\tilde{\varepsilon}$ is one draw from $N(0, \tilde{\sigma}^2)$(Tomz, Wittenberg, and King 2003, 26). It is dynamic in the sense that it incorporates a lagged dependent variable in each iterative estimation. At each iteration, the predicted value of the dependent variable given the scenario specified ($\tilde{y}_C|\boldsymbol{X_C}$) is used as the value of the lagged dependent variable (y_{t-1}) for the next iteration (to calculate $\tilde{y}_t$). In order to view the impact of key variables on the long-term dynamics, the series must be autoregressive, as shown by a statistically significant lagged dependent variable.[5] While few researchers make any attempt to present a dynamic interpretation of their findings, as Table 1 indicates, those who do seldom accompany their dynamic interpretations with confidence intervals or other appropriate indications of the estimated uncertainty surrounding their point estimates. In recent years, presenting some indication of uncertainty for substantive inferences has become standard in most areas of political science research. We see no reason why dynamic inferences should be an exception to this norm.

Our command uses King, Tomz, and Wittenberg's (2000) Clarify statistical package to present long-term dynamics for autoregressive relationships of continuous dependent variables. For a user-specified number of iterations, dynsim uses the n draws from the asymptotic sampling distribution with mean equal to the point estimates of the parameters and variance equal to the variance-covariance matrix of estimates (Clarify's estsimp) and the real or hypothetical values for the user-specified scenario, $\boldsymbol{X}_C$ (Clarify's setx), to simulate n expected values of the dependent variable (Clarify's simqi) (Williams and Whitten 2011).[6]

Studies like Rueda's would benefit from graphical depictions of substantively interesting dynamic simulations. For instance, it would be interesting to look at the impact of left-leaning versus right-leaning governments on employment policies over longer periods of time. Dynamic simulations allow researchers to make much richer and more nuanced inferences about dynamic relationships. Since the figures contain confidence intervals, it is possible to determine which relationships produce statistically significant effects in both the short- and long-term. Using this indication of uncertainty, we can then make additional inferences about whether scenarios have statistically different expected values at time t, and whether the expected value is statistically different over time. Not only can long-term dynamic simulations show the effects of key variables, but they also can illustrate how the predicted values respond to exogenous shocks and changes in the values of

[4]Our supplementary materials document contains two pieces of evidence to this effect. First, we provide a replication of Beck and Katz's classic Monte Carlo experiments using Clarify and analytical methods of estimation. The results across these two methods of estimation are almost indistinguishable. And, second, we provide a set of the same forecasts using dynamic simulations (built out of Clarify) and analytical methods. The confidence intervals for these pairs of forecasts were so close to each other that we had to use a statistical jittering technique to show them on the same figure.

[5]We warn against using any sort of dynamic simulation on an integrated series, as shown by a coefficient for the lagged dependent variable in which we cannot reject the null that the coefficient is equal to 1. Beck and Katz (2004) point out that researchers should not use stationary methods to analyze nonstationary data, as the causal inferences are likely to be grossly misleading. We therefore suggest ensuring that all series are stationary before estimating any dynamic simulations.

[6]Please see Williams and Whitten (2011) for a lengthy explanation of the details of how this command works. The Stata ado file can be downloaded at "web.missouri.edu~williamslaro/."

interactive relationships. We suggest that scholars fully explore these long-term effects in dynamic relationships so that they can make the full slate of inferences but also so that they can avoid making inferences that are only valid when examining effects in the short-term.

To illustrate our point, we provide two applied examples in which we are able to obtain additional substantive inferences by pushing our interpretation of TSCS models beyond simple point estimates and t-tests for short-term effects.

Two Illustrations

In this section, we illustrate the usefulness of presenting figures that depict the long-term dynamics of relationships. Modeling long-term dynamics allows for a much richer interpretation of the substantive effects of independent variables on the processes under examination. We begin with an example from an article by Poe and Tate (1994) in which they tested a series of theories about the determinants of state repression using data from 153 countries measured for eight consecutive years.

Example 1: Poe and Tate (1994)

The authors build a model of state repression based on a number of hypothesized relationships. The level of state repression should be lower in nations with higher levels of democracy, higher levels of economic development, and with a British cultural influence. Other factors should lead to higher levels of state repression: nations experiencing rapid economic growth, nations with rapid population growth, and nations under leftist or military regimes. Internal or external instability, characterized by civil and international war, is hypothesized to lead to more state repression. They test their hypotheses with an OLS estimation with White's robust standard errors and include a lagged dependent variable in the model to help correct for serially correlated errors (Beck and Katz 1996).[7]

In Table 2, we present the replicated results for their model with state repression operationalized as Amnesty International's personal integrity rights and democracy measured with Freedom House's political rights index (Poe and Tate 1994; 861, Model 1, Table 1).

Parameter estimates for the lagged dependent variable, level of democracy, population size, economic standing, and civil and international wars are statistically significant in the expected directions. Poe and Tate (1994) also provide graphical representations of the statistically significant variables of interest for the four estimated models. Their Figures 1–4 (in Poe and Tate 1994; 862–65) show the estimated changes in the predicted repression score over 10 years as a result of a loss of democracy, increase in economic standing, and presence of international war and civil war.

The figures that Poe and Tate (1994) present are helpful in that they present postestimation interpretations beyond what we see in most of the recent literature. The figures, however, lack measures of uncertainty, so it is impossible to know which relationships are statistically significant. Our postestimation interpretation procedures can present effectively the long-range dynamics of multiple scenarios and thus improve the causal inferences drawn from this research. When this is done, we can make two types of substantive inferences. First, we can note whether the two scenarios are significantly different from each other at any time period, and second, whether the predicted repression score for a given scenario at time *t* is statistically different from its score at any other time period.

Figure 1 presents the predicted Amnesty International human rights score (and 95% confidence intervals represented by the bars) over eight years for three different scenarios based on the same independent variables but holding economic standing at its minimum, mean and maximum, respectively.[8]

This type of figure has at least two advantages over graphs of the point estimates. First, placing 95% confidence intervals around the prediction substantially improves our inferential power. By comparing the simulated 95% confidence intervals for our different scenarios in the same year, we can now see that a country with the maximum level of economic standing has a level of human rights abuses that is statistically indistinguishable from the other two

[7]Several different authors have pointed out that many applied researchers have wrongly inferred from Beck and Katz's seminal article that a lagged dependent variable is the magical cure for all dynamic problems (e.g., Beck 2007; Beck and Katz 2004; Kittel and Winner 2005).

[8]The values of the other independent variables are held at either their means (for continuous variables) or their modes (for binary variables). Democracy, population size, change in population, and change in GNP per capita are held at their sample means, while left government, military control, British cultural influence, civil war, and international wars are held at their sample modes (0 in all cases).

TABLE 2 Replication Results for the Determinants of State Repression (Model 1, Poe and Tate 1994; 861)

Variable	Coefficient	(Std. Err.)	[90% CI]
Personal integrity abuse$_{t-1}$	0.730**	(0.024)	[0.691, 0.768]
Democracy	-0.045**	(0.013)	[-0.066, -0.024]
Population size	0.053**	(0.010)	[0.037, 0.069]
Population change	0.008	(0.012)	[-0.012, 0.028]
Economic standing	-0.008*	(0.003)	[-0.013, -0.002]
Economic growth	-0.001	(0.001)	[-0.003, 0.001]
Leftist government	-0.035	(0.054)	[-0.123, 0.054]
Military control	0.046	(0.047)	[-0.032, 0.123]
British cultural influence	-0.030	(0.040)	[-0.095, 0.036]
International war	0.208**	(0.079)	[0.077, 0.338]
Civil war	0.327**	(0.077)	[0.200, 0.454]
Constant	-0.021	(0.162)	[-0.288, 0.246]
R^2		.77	
Total N		1071	
Cross-sections		153	
Time Points		7	

*p $<$ 0.05 (two-tailed test); **p $<$ 0.01 (two-tailed test).
Note: Cell entries indicate OLS coefficients with robust standard errors in parentheses and 90% confidence intervals in brackets.

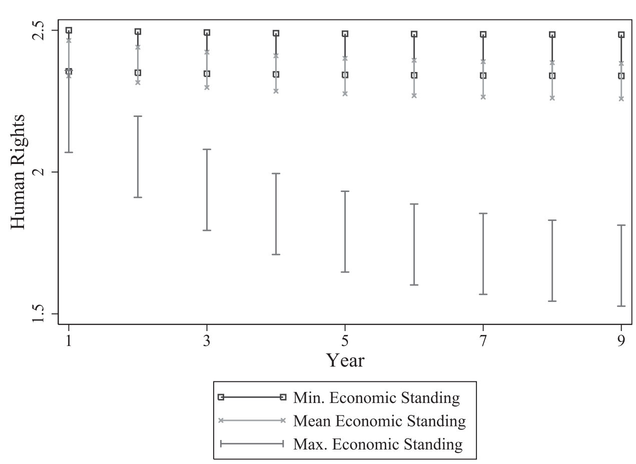
cases (mean and minimum economic standing) in year 1[9]. In year 2, this difference becomes statistically significantly lower and continues to decrease throughout the simulated time period. It is also apparent that by year 4 the level of human rights abuses in the country with the maximum level of economic standing is statistically different from what it was in this same nation during year 1.

Dynamic simulations may also help researchers to avoid making mistakes of interpretation. Figure 1 leads to a substantively different conclusion than the one reached by Poe and Tate (1994). Rather than assuming that an increase in economic standing produces the same beneficial response at all levels of economic standing, our figure shows that repression scores are only statistically lower for the states with the highest level of economic standing. In other words, out of the three scenarios of economic standing, the only scenario that is statistically different from the others in the long-run is the scenario with the maximum level of economic standing. This is the case throughout the last seven years of the simulation.

Poe and Tate argue that both civil and international wars increase the level of repression in a state. These findings are corroborated by the figures in their paper (their Figures 3 and 4), showing the change in

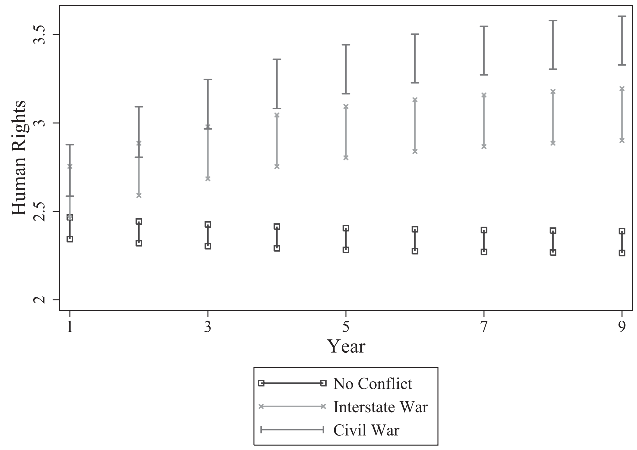
[9]It is worth noting that even though an independent variable may have a statistically significant estimated effect on the dependent variable, such effects do not always generate correspondingly statistically significant differences between predicted values. We see an example of this when we compare the results in Table 2 with those presented in Figure 1. The difference between two otherwise identical nations where one has the mean value for economic standing (3.516) and one has the minimum value (0.087) is clearly statistically significant (since we can see from Table 2 that a one unit shift in this variable is statistically significant). But the two 95% confidence intervals for these otherwise identical scenarios clearly overlap in Figure 1. This is the case because the variance for an OLS parameter estimate is calculated as $var(\hat{\beta}) = \frac{\sigma^2}{\sum(X_i - \bar{X})^2}$ while the variance for the predicted value of the dependent variable is calculated as $var(Y_0|\mathbf{x}_0 = \sigma^2[1 + \mathbf{x}_0^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{x}_0]$, where x_0 is the specified values of the *X*s (see, for example Gujarati 2003; 941–42). Assessments of statistical significance using predicted values of the dependent variable will thus always be more conservative.

FIGURE 1 Dynamic Simulations of the Effects of Economic Standing on Personal Integrity Rights



repression scores during times of civil and international war. When we model the long-term dynamics of the effects of interstate and civil wars on repression, we can make additional inferences regarding this relationship. Figure 2 shows the predicted repression score (and 95% confidence intervals) for three cases that are identical in year 0. These cases diverge in year 1 with one case having no conflict, one case experiencing a civil war, and one case embroiled in an interstate war.[10] In year 1, the case with no conflict has a statistically lower level of repression than the the case undergoing a civil war. Although their point estimates differ as expected, none of the other pairwise comparisons of cases are statistically distinguishable. In year 2, the case with no conflict is statistically lower than both cases with conflict, but the two conflict cases cannot be statistically separated. In year 4, we see that the two cases with conflict are barely statistically distinguishable with repression being higher in the case with a civil war than in the case experiencing interstate conflict.

FIGURE 2 **Dynamic Simulations of the Effects of Civil and Interstate Wars on Personal Integrity Rights**



Example 2: Whitten and Williams (2011)

Our next illustration comes from research on the political economy of defense spending (Whitten and Williams 2011). The authors begin with the observation that OECD democracies have been in a state of relative international security in the post-WWII period. Without the presence of international conflicts that threaten a state's existence, partisan actors have greater leeway to use defense spending to meet the needs of their domestic constituents. Rather than the traditional viewpoint of "guns versus butter," they suggest that, based on the empirical evidence of the relationship between military spending and economic indicators, this cliché should be modified to "guns *yield* butter." As a result, they expect that partisan leaders will use the low-level international conflicts that have characterized this era as opportunities to alter defense spending to satisfy their domestic constituencies. Rather than assuming that there is only one relevant dimension of partisan preferences (right versus left) that influences defense spending, they show that two ideological dimensions—a government's welfare position and its international peace position—have impacts on defense spending. Governments that favor hawkish positions or more generous welfare spending will have higher levels of defense spending than more dovish or austere governments. They find that these partisan effects appear even while controlling for the state of the economy (*Real GDP Growth*$_{t\text{-}1}$), the domestic political conditions (*Minority Government*, *Number of Government Parties*, and *Election Year*), the international strategic environment (*Alliances*$_{t\text{-}1}$, *US/Soviet CINC Ratio*$_{t\text{-}1}$, *and Changes in US Military Expenditures as a Percentage of GDP*$_{t\text{-}1}$), and capabilities (*CINC Score*$_{t\text{-}1}$). The numerical results are shown in Table 3.

To illustrate the effects of these two dimensions of government ideology on defense spending, we predict the level of defense spending for four governments in the corners of our two-dimensional distribution (5th and 95th percentiles) of ideological positions (austere-hawks, austere-doves, generous-hawks, and generous-doves). Figure 3 shows how these four types of governments respond to an external shock in the form of large changes in U.S. defense spending. They argue that changes in U.S. defense spending represent signals that OECD democracies also should increase their defense spending. Whether this is in anticipation of a future threat or simply taken as a proxy for the state of the Cold War, Whitten and Williams (2011) anticipate that the level of defense spending will increase for each type of government. This appears to be the case; an increase of over 1.5% of the U.S. defense budget in the late

[10]The "no conflict" scenario shows the repression score when all of the continuous variables are at their sample means and the binary variables are at their modes (0 for all variables). In the civil war scenario, the variable for civil war is set at 1. The final scenario has the value for international war set to 1. It is not necessary to restrict the variables to their sample means and modes; in fact, richer inferences may be made by placing the values of the variables at other substantively important values. For example, we could make the economic conditions mirror typical values for war-torn states or some set of values from a historical case.

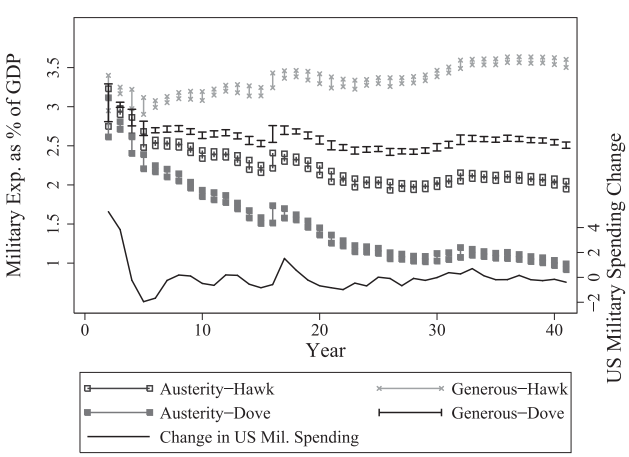
TABLE 3 Two Dimensional Model of Government Ideology and Defense Spending (Whitten and Williams 2011)

Variable	Coefficient	(Std. Err.)	[90% CI]
Military expenditures as a % of GDP_{t-1}	0.933**	(0.020)	[0.901, 0.965]
Minority government	0.033	(0.031)	[-0.019, 0.084]
Number of government parties	0.008	(0.010)	[-0.009, 0.025]
Election year	0.008	(0.027)	[-0.035, 0.052]
Real growth in GDP_{t-1}	0.404	(0.554)	[-0.508, 1.315]
CINC $score_{t-1}$	2.231	(2.289)	[-1.535, 5.996]
$Alliance_{t-1}$	0.025	(0.031)	[-0.032, 0.082]
US change in mil. exp. as a % of GDP_{t-1}	0.081*	(0.037)	[0.020, 0.143]
US/Soviet CINC $ratio_{t-1}$	-0.026	(0.061)	[-0.127, 0.075]
Conflict involvement (MIDs composite)	0.012*	(0.006)	[0.002, 0.021]
Government welfare position	0.007**	(0.003)	[0.002, 0.011]
Government welfare position × conflict	-0.001	(0.001)	[-0.001, 0.000]
Government hawk position	0.009	(0.006)	[-0.001, 0.019]
Government hawk position × conflict	0.001	(0.002)	[-0.001, 0.004]
Constant	0.047	(0.090)	[-0.101, 0.196]
R^2		.92	
Total N		776	
Cross-sections		19	
Time Points		19 to 46	

*$p < 0.05$ (two-tailed test); **$p < 0.01$ (two-tailed test).
Note: Cell entries indicate OLS coefficients with robust standard errors in parentheses and 90% confidence intervals in brackets.

1960s (around year 16) leads to an immediate jump in defense spending for all four governments. In all four cases, this increase is an important one, as the predicted level of spending is statistically higher (or nearly so) in the next period. These types of figures also illustrate how long it takes various governments to revert to their previous level of spending. For generous-hawks, who have the highest level of spending, the level of spending never returns to its preexogenous shock level but continues to climb throughout the simulation. Generous-doves and austerity-hawks, the next two highest levels, take approximately five to six years to revert to the original level. The effects of exogenous shocks for austerity-doves, on the other hand, are not statistically different after the shock; their levels of spending return to the preshock level (and actually become statistically lower) much faster than the other three governments.

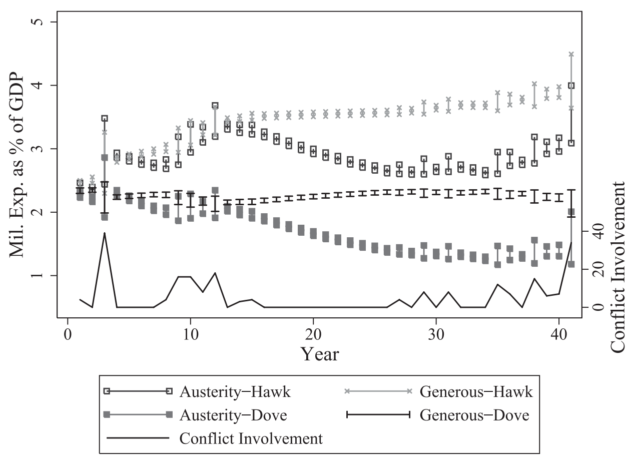
FIGURE 3 Predicted Defense Spending by Four Hypothetical Governments as they Respond to U.S. Changes in Defense Spending



Note: bars depict 95% confidence intervals

As three excellent recent works on interactions emphasize (Braumoeller 2004, Brambor, Clark and Golder 2006; Kam and Franzese 2007), graphical illustrations are crucial to the appropriate interpretation of interactive relationships. This is especially true when the interactive relationships are estimated over time. Figure 4 presents the predicted level of military spending as a percentage of GDP (and their 95% confidence intervals) for four hypothetical governments (based on the two dimensions of welfare and international peace) across the international conflict levels that France faced from 1950 to 1989. The only values that change at each iteration are the values of the lagged military spending as a percentage of GDP and the interaction of the government welfare and hawk positions (which become the products of each of the

FIGURE 4 Predicted Defense Spending by Four Hypothetical French Government Types as they Respond to Conflict Involvement



ideology variables and the conflict involvement variable). At each iteration, the value of the conflict involvement variable becomes the concurrent value of France's conflict involvement for that time period.

This figure suggests that governments react to conflict involvement differently, depending on their positions on welfare. During times of conflict, hawks (both austere and generous) have higher levels of spending than doves (both austere and generous), but there is no difference between governments based on welfare. However, when there is no conflict (between year 15 and 28), the differences between austere and generous governments become statistically different from one another because of the drastic cuts in defense spending orchestrated by austere governments. This is a counterintuitive finding, one that might have been ignored without a graphical representation of the long-term dynamics of the interaction between ideology and international threat.

In the previous four figures, we have shown the utility of long-term dynamic simulations with the aid of two research applications. Not only can long-term dynamic simulations show the effects of key variables, but they also can illustrate how the predicted values react to exogenous shocks and changes in the values of interactive relationships.

Conclusion

We hope that this article leads to an increase in the use of dynamic simulations for interpreting the results of models estimated with TSCS data. As the extensive technical literature in this area shows, this class of models is prone to many statistical pitfalls. Despite these problems, the use of these models is increasing, because they provide valuable and otherwise unavailable leverage to answer important questions about the political world. If researchers go through all the trouble to obtain believable estimates with TSCS models, they should take the time to make correct and useful inferences from their results.

Acknowledgments

Earlier versions of this article were presented to audiences at the annual meeting of the Midwest Political Science Association, Rice University, the University of Houston, the Summer School in Concepts, Methods and Techniques in Political Science at the University of São Paolo, the St. Louis Area Methods meeting, and Concordia School in Social Science Data Analysis (Montreal). The authors thank members of these audiences and, in particular, Luke Keele, Tim Hellwig, Jeff Gill, Harold Clarke, Robert Walker, Patrick Brandt, Jamie Monogan, and Harvey Palmer for their helpful comments. Despite this wealth of helpful advice, we remain responsible for all errors in this article. The replication files and online appendix are available at web.missouri.edu/~williamslaro/.

References

Beck, Nathaniel. 2007. "From Statistical Nuisance to Serious Modeling: Changing How We Think About the Analysis of Time-Series–Cross-Section Data." *Political Analysis* 15 (2): 1–4.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data." *American Political Science Review* 89 (3): 634–47.

Beck, Nathaniel, and Jonathan N. Katz. 1996. "Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models." *Political Analysis* 6 (1): 1–36.

Beck, Nathaniel, and Jonathan N. Katz. 2004. "Time-Series–Cross-Section Issues: Dynamics, 2004." Paper presented at the 2004 annual meeting of the Society of Political Methodology, Stanford University.

Bewley, R. A. 1979. "The Direct Estimation of the Equilibrium Response in a Linear Model." *Economic Letters* 3: 357–61.

Brambor, Thomas, William Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1): 63–82.

Braumoeller, Bear F. 2004. "Hypothesis Testing and Muliplicative Interaction Terms." *International Organization* 58 (4): 807–20.

deBoef, Suzanna, and Luke Keele. 2008. "Taking Time Seriously: Dynamic Regression." *American Journal of Political Science* 52 (1): 184–200.

Enders, Walter. 2004. *Applied Econometric Time Series*. 2nd ed. India: Wiley.

Freeman, John R., John T. Williams, and Tse min Lin. 1989. "Vector-Autoregression and the Study of Politics." *American Journal of Political Science* 33 (4): 842–77.

Green, Donald P., Soo Yeon Kim, and David H. Yoon. 2001. "Dirty Pool." *International Organization* 55 (2): 441–68.

Greene, William H. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Pearson.

Gujarati, Damodar. 2003. *Basic Econometrics*. New York: Mc-Graw Hill.

Kam, Cindy D., and Robert J. Franzese, Jr. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis: A Refresher and Some Practical Advice*. Ann Arbor: University of Michigan Press.

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): 347–61.

Kittel, Bernhard, and Hannes Winner. 2005. "How Reliable Is Pooled Analysis in Political Economy? The Globalization-Welfare State Nexus Revisited." *European Journal of Political Research* 44 (2): 269–93.

Poe, Steven C, and C. Neal Tate. 1994. "Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis." *American Political Science Review* 88 (4): 853–72.

Rueda, David. 2005. "Insider–Outsider Politics in Industrialized Democracies: The Challenge to Social Democratic Parties." *American Political Science Review* 99 (1): 61–74.

Stimson, James A. 1985. "Regression Models in Space and Time: A Statistical Essay." *American Journal of Political Science* 29 (4): 914–47.

Tomz, Michael, Jason Wittenberg, and Gary King. 2003. "CLARIFY: Software for Interpreting and Presenting Statistical Results." *Journal of Statistical Software* 8 (1): 1–30.

Whitten, Guy D., and Laron K. Williams. 2011. "Buttery Guns and Welfare Hawks: The Politics of Defense Spending in Advanced Industrial Democracies." *American Journal of Political Science* 55 (1): 117–35.

Williams, Laron K., and Guy D. Whitten. 2011. "Dynamic Simulations of Autoregressive Relationships." *The Stata Journal* 11 (4): 1–12.

Laron K. Williams is Assistant Professor in the Department of Political Science at the University of Missouri, Columbia, MO 65211-6030.

Guy D. Whitten is Associate Professor in the Department of Political Science at Texas A&M University, TAMUS 4348, College Station, TX 77843.