Coversheet

This is the accepted manuscript (post-print version) of the article.
Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

Publication metadata

A Dynamic Linear Modeling Approach to Public Policy Change

Matt W. Loftis*
Aarhus University
mattwloftis@ps.au.dk

Peter B. Mortensen
Aarhus University
peter@ps.au.dk

December 2018

Abstract

Theories of public policy change, despite their differences, converge on one point of strong agreement. The relationship between policy and its causes can and does change over time. This consensus yields numerous empirical implications, but our standard analytical tools are inadequate for testing them. As a result, the dynamic and transformative relationships predicted by policy theories have been left largely unexplored in time-series analysis of public policy. This paper introduces dynamic linear modeling (DLM) as a useful statistical tool for exploring time-varying relationships in public policy. The paper offers a detailed exposition of the DLM approach and illustrates its usefulness with a time series analysis of U.S. defense policy from 1957-2010. The results point the way for a new attention to dynamics in the policy process and the paper concludes with a discussion of how this research program can profit from applying DLMs.

Published: Journal of Public Policy

Many policy theories incorporate the idea that, in one way or the other, the causal effects driving public policy change over time. The Advocacy Coalition Framework proposed by Sabatier (1987, 1988) and later developed in collaboration with Jenkins-Smith (Sabatier & Jenkins-Smith 1993, Jenkins-Smith, Nohrstedt, Weible & Sabatier 2014) is one. A second is the theory of social learning and policy paradigms introduced by Hall (1993), and a third theory is the Punctuated Equilibrium Theory developed by Baumgartner & Jones (1993).

As laid out in this paper, the three theoretical approaches differ in their conceptualization, description, and explanation of policy change. However, they all depict a policy-making process in which decision-makers' weighting and evaluation of what is important to policy decisions is likely to change over time (for reviews, see Weible, Sabatier & McQueen 2009, Cairney & Heikkila 2014, Baumgartner, Jones & Mortensen 2014, Hogan & Howlett 2015). Thus, policy theory may tell us that an effect waxes or wanes from one period to another, or even that it changes direction. Such change may be slow or fast, gradual or punctuated. It may reflect change in policy makers' core beliefs, change in the paradigms dominating policymaking, or change in the images of various policy areas. Nevertheless, a common implication of the dominant public policy theories is that the importance of explanatory factors varies over time.

In this paper, we propose a dynamic linear modeling approach as a means of more closely aligning our statistical modeling choices with the dynamic implications of the dominant theories of public policy. Some policy studies have attempted to take into account time-varying effects by specifying interaction terms or by making *a priori* predictions about the timing of shifts and accordingly dividing time series into various epochs. Such modeling approaches that address temporal heterogeneity in coefficient estimates come closer to the ontology of our theories of public policy, but they overlook an important theoretical gap: policy theories rarely make point predictions that tell us when changes in the relationship between x and y take place. Furthermore, even if a theoretical argument identifies a specific point in time when parameters are expected to shift, it requires a flexible estimation method

2

to evaluate such a claim empirically. Thus, we argue that public policy researchers should favor estimation methods that are guided by the data and their chronology, rather than *a priori* information, with regard to the timing and direction of shifts.

The dynamic linear modeling approach is aimed at directly estimating time-varying influences on public policy. The method is familiar in political science, but has primarily been used to study public opinion and international conflict (Beck 1983, Gerber & Green 1998, Green, Gerber & de Boef 1999, Mitchell, Gates & Hegre 1999, Brandt, Williams, Fordham & Pollins 2000, Wood 2000, Mcavoy 2006, Enns & McAvoy 2011, Fariss 2014). There are other approaches to identify time-varying relationships, but as we argue, the DLM approach is particularly well-aligned with the challenges posed by dynamic public policy theories.

In this paper, we describe and demonstrate the basic methodology for unfamiliar readers and argue that the tools currently available for building and estimating these models allow researchers to build the right models for their own policy data. Fundamentally, we argue that, with regard to investigating dynamic effects on public policy, the method we propose accomplishes this efficiently and in a way that approximates the processes described by the major theories of public policy.

To corroborate this claim, we conduct a dynamic linear model analysis of change in U.S. defense policy from 1956-2010. The case of U.S. defense is a particularly attractive test ground given the many previous time-series studies of this policy area, which provides direction for the choice of explanatory variables, and given the availability of high-quality data spanning several decades.

Motivating a dynamic approach to modeling public policy

A growing body of public policy theories argues that politics is a terrain that is restructured over time. A prime example is the work of Baumgartner, Jones, and colleagues, which claims that over time most policy areas may be subject to changes in policymakers' focus (Jones 1994, Baumgartner & Jones 2002, Jones & Baumgartner 2005, Baumgartner & Jones 2015). As a consequence, no single logic of policy making and no single frame of reference lasts forever. Historical contingencies are fundamental to politics, and the threshold for action even within a single policy area may change over time, as the weighting of policy problems is dynamic (Jones & Baumgartner 2005, 90-91).

Other major contributions to the study of public policy have a different conceptualization, description, and explanation of policy change, but they share the important implication that the determinants of public policy may change over time. One example is the Advocacy Coalition Framework, according to which every coalition of policy makers contains secondary aspects, policy core beliefs, and so called deep core beliefs. Secondary aspects comprise a large set of narrower beliefs of the policy makers concerning, for instance, their evaluation of relevant problems and performances within a policy subsystem as well as their preferences regarding regulations or budgetary allocations (Sabatier 1998). Policy core beliefs refer to a coalition's basic normative commitments and causal perceptions across an entire policy domain, and deep core beliefs include basic ontological and normative beliefs operating across almost all policy domains.[1] When it comes to the core beliefs, these are assumed to be very stable, though not time-invariant. Even if changes mainly involve secondary aspects, this may, for instance, have implications for the relative importance of various causal factors within a given policy domain (Sabatier 1998, 104).

Another example is Hall's (1993) seminal work on social learning and policy paradigms, according to which both the instruments and goals of public policymaking may change over

time (see also Campbell 2002, Blyth 2013, Princen & 't Hart 2014). In Hall's thinking, policy change comes in three basic types. First-order changes are policy shifts that leave the larger policy regime in place, for instance, changing the minimum lending rate or fiscal stance to adjust macroeconomic policy. In Hall's analogy, first-order changes adjust the settings on the "instruments" of policy making. Second-order changes, on the other hand, replace the instruments themselves (Hall 1993, 278-9) – in the example of macroeconomic policy, changes in the system of monetary control or adopting or abandoning strict targets for monetary growth would be second-order changes. Third-order changes alter the instruments, their settings, and the hierarchy of goals behind policy. Hall's prime example of a third-order change is the paradigmatic shift in British macroeconomic policy in the 1970s and 1980s from Keynesian to monetarist modes of economic regulation.

Just as changes in deep core beliefs are assumed to be rare according to the ACF, paradigmatic third-order changes are very rare according to Hall's theory. It is important to note, however, that even second-order changes or, in the parlance of ACF, changes in policy core beliefs will most likely be reflected in time-varying relationships in time-series models of public policy. A second-order change, for instance, reflects an alternation of the hierarchy of goals behind policy (Hall 1993, 282), and a change in the policy core beliefs of policy makers may involve a change of their causal perceptions of a policy domain (Sabatier 1998, 103). Put differently, if the decision makers' causal understanding of the relationship between problems and solutions changes within policy domains such as defense, welfare, transportation, or environmental policies, then the relationship between x and y probably changes as well.[2]

Across the different theoretical approaches and different subfields of policy research there is a consensus that the dynamic implications of these theories are not adequately analyzed by using standard time series regression techniques (Hall 2006, Jones & Baumgartner 2005, 90-91). The basic problem with standard regression models is simple. If relationships between x and y are time-variant then the basic regression assumption of unit homogeneity is violated. In other words, changes in the value of an explanatory variable, x, will not produce

corresponding changes in the value of the policy output variable, y, of the same magnitude, or even direction, across all time (cf. Hall 2003, 382).

As we will discuss in detail below, common responses to this problem are to divide time series into epochs or to specify interaction terms conditioning the relationship between x and y. The problem, however, is that such approaches require strong theories about when the relationship between x and y transforms. Often, as argued by Jones & Baumgartner (2005, 23), we have no *a priori* idea when such changes take place: "If we look at a series of policy activities, as we do for crime, social welfare, and other policy arenas, we find clear evidence of self-reinforcing changes. While it is possible to model such complex series, how does one know in advance when these interactions will occur?" Similarly, in a recent review of the Advocacy Coalition Framework, Jenkins-Smith et al. (2014) identify four major pathways to policy change, each of which may be a necessary but not sufficient source of change in policy core beliefs. Finally, even if theory does contain a strong prediction of the timing of a transformative policy change, we need a flexible estimation model in order to be able to falsify or support that prediction empirically.

This discontent with standard time series approaches has led policy scholars to advocate alternative methods that are better aligned with their theoretical approaches to public policy. One example is systematic process tracing (Hall 2003), which can be a strong approach to identifying causal processes leading to a political outcome. Another example is the stochastic process approach advocated by Jones & Baumgartner (2005). A major advantage of a stochastic process approach is the focus on entire distributions of data (not only mean and variance) and the lack of assumptions required. Yet, we also set aside important information by using these approaches: giving up the ability to examine directly the relationships between changes in (potentially multiple) policy determinants and changes in policy outputs or ignoring the chronology of our data. Thus, we sympathize with the expansion of methods better aligned with the ontology of the major public policy theories, but we believe there also is a strong case for improving the statistical time series analysis of policy dynamics.

In the next section, we motivate the adoption of a flexible approach to time-series policy analysis suited to uncovering time-varying relationships.

Estimating dynamic relationships in public policy research

Given the goal of examining dynamic relationships between an explanatory variable (x) and a policy output (y), we propose that the hypothetical ideal statistical model has two features. First, it would estimate variation in x's relationship to y over time (t), directly from the data. Formally, we might write:

$$y_t = x_t \beta_t + \epsilon_t \tag{1}$$

The t subscript on β_t indicates a new parameter value is estimated each time period, indicating how the relationship between x_t and y_t changes or remains constant. Such a model's estimates would be guided by the data, since they would be estimated directly.

Secondly, an ideal model would estimate β_t in a way that respects its chronology. That is, given the emphasis theory places on history and path dependence in public policy (cf. Jones & Baumgartner 2005, 49), we must not assume $\beta_{t'}$ is completely independent of $\beta_{t''}$ (e.g. β_t is not independent of β_{t-1} or of β_{t-4}, etc.).

The main barrier to this idealized model is identification — that is, its parameters outnumber its data points and so, even with infinitely long time series, one cannot estimate them without making additional assumptions. The time-series regression modeling toolkit contains many second-best alternatives, however. We argue that time-varying relationships of the form presented in equation 1 are common in policy studies, but that the tools commonly used to study them provide answers that are not guided only by the data nor do they account for the data's chronology.

In this section, we provide an accessible introduction to a well-studied method for estimat-

ing dynamic relationships that meets both criteria; we review the advantages these criteria hold for public policy research; and we consider the method's strengths relative to commonly used approaches. Namely, we propose that researchers examining policy changes featuring time-varying relationships of this sort consider a class of modeling techniques known as dynamic linear models (DLM) (Kim & Nelson 2000, Mcavoy 2006, Shumway & Stoffer 2010) or flexible least squares (Kalaba & Tesfatsion 1988, Wood 2000).[3] The difference between the two, for our purposes, lies assumptions about the distribution of errors in DLMs. We refer to DLMs throughout since we adopt this formulation to estimate uncertainty.

In what follows, we make the case for more widely adopting DLMs for studying policy change. It is worth reiterating that DLMs, a subset of the broader category of state-space models, are not unknown in political science (Gerber & Green 1998, Green, Gerber & de Boef 1999, Mitchell, Gates & Hegre 1999, Brandt et al. 2000, Wood 2000, Mcavoy 2006, Enns & McAvoy 2011, Fariss 2014). Indeed, Beck (1983) argued that the approach had strong advantages for theory testing and should be added to the arsenal of those who work with time series data. Though they have been applied to study public opinion and international conflict, specifically to deal with the kind of effect heterogeneity that we are concerned with, the method has not gained traction in studies of public policy. Potential applications are, nevertheless, widespread. The goal of estimating time-varying relationships originally motivated state-space models' development in physics and, more recently, has encouraged their adoption and popularization especially in economics and finance (for background, see: Kim & Nelson 2000, Mergner 2009).

Returning to equation 1's hypothetical model, the DLM recovers estimates of time-varying effect coefficients, i.e. β_t, provided some scaffolding in the way of assumptions about how β_t changes with time. Specifically, in a DLM β_t is a realization of an underlying "state" that varies as a Markov chain over time. This formulation ensures identification and reflects one of the principles we proposed above: that dynamic relationships should be estimated in a way that respects the data's chronology. As we will see, DLMs can flexibly return

time-constant coefficient estimates where appropriate or detect the presence of dynamics. A simple DLM built to emulate equation 1 might be structured as follows:

$$
\begin{aligned}
y_t &= \boldsymbol{X_t}\boldsymbol{\beta_t} + v_t & (2)\\
v_t &\sim i.i.d.(0, \sigma_v^2)\\
\boldsymbol{\beta_t} &= \boldsymbol{\beta_{t-1}} + \boldsymbol{w_t} & (3)\\
\boldsymbol{w_t} &\sim i.i.d.(0, Q)
\end{aligned}
$$

The first line, equation 2, is called the measurement equation. Inputs to the policy-making process in vector $\boldsymbol{X_t}$ influence policy y_t via the vector of time-varying effect coefficients $\boldsymbol{\beta_t}$. The disturbances, v_t, are independent zero-mean Gaussian noise with variance σ_v^2. The third line, equation 3, is the state equation. Here, the effect coefficients, $\boldsymbol{\beta_t}$, are modeled as an unobserved state varying in a random walk over time. The state disturbances, in vector $\boldsymbol{w_t}$, are also mean-zero Gaussian noise with covariance matrix Q and are uncorrelated to v_t.

Equations 2 and 3 demonstrate how a DLM respects the goal of producing estimates guided by the data and accounting for chronology. Estimates of $\boldsymbol{\beta_t}$ come directly from the data and are linked over time in equation 3. This structure allows estimates of $\boldsymbol{\beta_t}$ to be informed by the whole time series, while still ensuring them flexibility to reflect any pattern of variation indicated by the data. What value added does this bring to policy studies examining time-varying relationships? We suggest several. For one, estimates that depend only on the data limit the assumptions analysts must make. As we review below, many methods for identifying dynamic relationships must justify not only their measurements, cases, and modeling choices — as in any empirical analysis — but also must justify *a priori* expectations about the timing of shifts in relationships or the use of statistical tests to pinpoint them.

Perhaps more importantly, estimates of dynamic relationships arising from the data enrich our ability to falsify hypotheses. Hypotheses about time-varying relationships may be falsified if a relationship does or does not exist, does or does not vary, or does or does not vary at a particular time. DLMs can speak to these criteria simultaneously. This has the advantage of avoiding bias from or corrections for multiple testing if the analyst wishes to consider different null hypotheses. Furthermore, since DLMs feature time-connected estimates of $\boldsymbol{\beta_t}$, they utilize all the information in the data, maximizing the statistical power of the estimates and using information from the full time series to estimate each coefficient.

Consider, briefly, advantages DLMs enjoy relative to other approaches to estimating time-varying relationships. A common approach to this problem is to estimate variation in effect parameters by subdividing time series into shorter epochs and running separate regressions on each subset of the data or, similarly, including interaction terms for epochs. This encounters several weaknesses. The first relates to statistical power. Subdividing time series necessarily weakens inferences by working with smaller data subsamples, increasing risks of both Type I and Type II errors. Secondly, and potentially more challenging, subdividing requires analysts to identify and motivate the choice of relevant time periods *a priori*. A principled choice of time periods may be available for certain tests, but considering more than one choice ventures into multiple testing. Furthermore, subdividing still forces analysts to assume effect parameters are constant within epochs.

Low statistical power is also a difficulty for CUSUM and CUSUMSQ plots as tests for parameter stability (Wood 2000, Baltagi 2011). These plot functions of cumulative sums of residuals to identify shifts in parameters. Though this approach is estimated directly from the data, it falls short of the criterion of time-connectedness. Having identified a change in a parameter via CUSUM plotting, this method provides no guidance in estimating or accounting for temporal dependency in parameter shifts.

Difficulties in motivating *a priori* information also apply to the Chow test, which can reveal if an effect parameter shift occurs at a particular point (Beck 1983, Wood 2000) —

10

a point the analyst identifies. Again, having identified potential change points, there is no clear method for estimating parameter changes that respect time-dependence.

Concerns with statistical inefficiency and motivating *a priori* information also extend to moving window analysis (Kwon & Pontusson 2010, Finseraas & Vernby 2011). Moving window analyses ignore large portions of data to estimate each bit of dynamics in effect coefficients. Though eventually all data is considered, observations at the beginning and end of time series are less influential than those in the middle, and each individual estimation has less statistical power than would be achieved using all data. Furthermore, decisions such as window size and which observations to include in each analysis can be consequential for the results, yet are left to the analyst's choice.

Still more systematic methods include multivariate generalized autoregressive conditional heteroskedasticity models (MV-GARCH) or dynamic conditional correlations (DCC) (Engle 2002, Lebo & Box-Steffensmeier 2008). These dynamically estimate the covariance matrix of multivariate time series using a framework adapted from GARCH modeling. This is a powerful method for identifying a certain type of dynamics, namely in correlations. Though they satisfy our proposed criteria of being data-driven and (partially) time-connected, they require quite large data sets and address themselves to correlations and not to the explanatory modeling problem described in equation 1.[4]

There are, of course, limitations to DLMs. As in any parametric modeling, the assumptions underlying the method matter and model reliability is better confirmed than assumed. Therefore, we provide guidance in probing the conditions for applying DLMs by using appropriate diagnostics and demonstrate how to use in-sample predictions and simulations to check the appropriateness of DLMs for a given application.[5]

Estimating a DLM

A standard approach to estimating DLMs like the one we describe above is via a combination of maximum likelihood and the Kalman filter (Kim & Nelson 2000, Comman-

deur & Koopman 2007, Petris, Petrone & Campagnoli 2009, Shumway & Stoffer 2010), though Bayesian estimation methods are also common (Petris, Petrone & Campagnoli 2009, Shumway & Stoffer 2010). We describe here how to apply the former approach to estimate a DLM. To illustrate the method, we walk through the process of estimating the model described in equations 2 and 3 — an example of a dynamic model of the impact of various influences on some policy output in a single time series.

DLMs are estimated recursively. Referring to equation 3, at each time t, $\boldsymbol{\beta_{t-1}}$ serves as our expectation of the new period's value of $\boldsymbol{\beta_t}$, conditional on the information observed up to time $t-1$. Based on this conditional expectation for $\boldsymbol{\beta_t}$, we calculate a prediction for the outcome in time t: $\hat{y}_t$. The error in this prediction, $y_t - \hat{y}_t$, is used to update our final estimate of $\boldsymbol{\beta_t}$, such that larger errors provoke larger shifts in coefficient estimates. The recursion proceeds sequentially from time 1 through time n. This means that at any point, t, all past information about the underlying state is summarized in the point estimates, $\boldsymbol{\beta_t}$, and their covariance matrix at time $t-1$.

Shifts in the coefficient estimates each period are moderated by the ratio of uncertainty regarding the estimates of $\boldsymbol{\beta_t}$ to the total overall estimation uncertainty in the model (including that from $\boldsymbol{\beta_t}$). Where uncertainty about $\boldsymbol{\beta_t}$ is a larger share of overall model uncertainty, this ratio — termed Kalman gain — is closer to one. Updates to the coefficients become more responsive to that period's prediction error as the Kalman gain approaches one. A formal definition of the Kalman gain and detailed estimating equations are provided in appendix A.

The final step in the estimation is to smooth the time-varying state estimates using information from the full time series. Kalman smoothing utilizes the same approach as the filter, but run in reverse. Recall that when filtering at each time period, t, all information about the unobserved states up to period t is summarized in the point estimates and associated uncertainty of last period's state estimates: $\boldsymbol{\beta_{t-1}}$. The smoother, by running in reverse from time T back to time 1, updates each period's estimate of $\boldsymbol{\beta_t}$ conditional on all information in the model. Filtering alone is used for applications in which the goal is to forecast future

12

outcomes, whereas smoothing is applied in applications aimed at making inferences about effect parameters. In public policy, our inferences are generally of the latter type, addressing effect estimates in fixed time series. For that goal, filtering by itself suffers several weaknesses relative to filtering plus smoothing. First, changes in filtered parameter estimates lag changes in smoothed parameter estimates (Commandeur & Koopman 2007, pp. 85–89). This is a product of the estimation process and does not negatively impact filtering's predictive accuracy. However, the lagged filtered parameter estimates are out of sync with the best estimate of the actual pattern in the data — a situation smoothing resolves. Second, filtering overstates uncertainty in early estimates of $\boldsymbol{\beta_t}$, since chronologically earlier observations are estimated using fewer data points. Third, filtering can over- or understate the magnitude of shifts in coefficient estimates relative to smoothed estimates. In sum, smoothing provides the more accurate inference for $\boldsymbol{\beta_t}$ because it uses more information than filtering alone (Kim & Nelson 2000, p. 27). Furthermore, as we have argued our estimates of $\boldsymbol{\beta_t}$ must not be fully independent of one another, we prefer combined filtering and smoothing.[6]

Final effect coefficient estimates are these smoothed time-varying coefficients. The recursive estimation procedure is identified conditional on a set of starting values for $\boldsymbol{\beta}$ ($\boldsymbol{\beta_0}$) and estimates of Q, σ_v^2, and the starting value of the covariance matrix of innovations, Σ_0. These are estimated via maximum likelihood. The process begins with initial values for each of these latter parameters, and then each $\boldsymbol{\beta_t}$ is estimated from time one using the Kalman filter followed by smoothing.

As can be seen in equation 2, the DLM also relies on assumptions about the independence, normality, and homoskedasticity of errors. Violations of these assumptions can be checked straightforwardly by examining standardized residuals from an estimated model (see, for example, Commandeur & Koopman 2007, pp. 90–96). An assumption common to standard time-series techniques that is not encoded in the DLM framework, is stationarity. That is, our methods often require we assume that statistical features of time series inputs to the model, like their mean and variance, do not vary over time. Such variation cannot be captured by

static coefficient estimates. However, since coefficients vary with time in the DLM data used in such applications need not exhibit stationarity (Commandeur & Koopman 2007, p. 134).[7]

The basic DLM described here can also be extended to accommodate a variety of alternative settings, for example: autoregressive processes, moving averages, constraining certain parameters to be constant, generalized linear models for non-normally distributed response variables, and others (Petris, Petrone & Campagnoli 2009, Shumway & Stoffer 2010). As we show later, the DLM can also be extended to panel data.

Statistical software for estimating DLMs is widely available. An excellent starting point for applying them in the reader's own work is to review volume 41 of the *Journal of Statistical Software*, a special issue on state space estimation in STATA, R, SAS, RATS, and other statistical software packages. The results we present are estimated using the "dlm" library for R (Petris 2010), which is one of several options in R (Petris & Petrone 2011, Tusell 2011).

We turn now to applying DLMs to a real empirical case in policy studies to demonstrate its usefulness and potential applications.

A dynamic linear model analysis of U.S. defense policy

We illustrate the dynamic modeling approach using real data on U.S. defense spending, applying the technique to test for dynamic relationships between the explanatory variables and policy outputs. U.S. defense spending offers an attractive test ground for several reasons. First, though the use of public spending as a policy indicator has been subject to some debate, the defense area is one where changes in spending have often been taken as evidence of important changes in defense policies (see e.g. True 2002, Wlezien 1996, Hartley & Russett 1992). Second, many studies have tried to model time series of U.S. defense spending, which provides some direction for our choice of explanatory variables. The many defense studies partly reflect a genuine interest in this policy field during the Cold War, and partly reflect the availability of rather reliable and long time series within this policy field. A broad

range of drivers of U.S. defense spending have been examined in these time series studies including election cycles (Nincic & Cusack 1979, Cusack & Ward 1981, Zuk & Woodbury 1986, Kamlet & Mowery 1987), measures of unemployment (Griffin, Wallace & Devine 1982, Cusack 1992, Su, Kamlet & Mowery 1993, Majeski 1992, Kiewiet & McCubbins 1991), arms races (Ostrom 1977, Correa & Kim 1992, Ostrom 1978, Ostrom & Marra 1986), inertia (Ostrom 1977, Majeski 1983, Correa & Kim 1992), and public opinion (Hartley & Russett 1992, Wlezien 1996, Higgs & Kilduff 1993). A more complete overview of variables and measures adopted in previous work is included in table A.1 in appendix B.[8]

The U.S. defense spending studies disagree with respect to several model assumptions, including the choice of explanatory variables and the exact specification of the dependent variable. Nevertheless, common to virtually all of them is the modeling assumption that relationships are invariant over time. The assumption is so uncontested that it is not given serious consideration in most previous defense spending studies. This is a further argument for investigating whether the approach advocated in this paper is able to detect dynamic relationships within such a well-researched field as U.S. defense spending. Only one study we know of escapes this characterization of the field. True (2002) examined U.S. defense spending over several decades and found that the relationship between defense spending and prior spending, international tensions and wars, and U.S. intelligence estimates of Soviet defense spending showed some evidence of changing over time.

To reach this conclusion, True (2002) split the time series into two periods, 1966-1979 and 1980-1992. Given the limitations with this research strategy we reviewed above, we re-evaluate True's findings using a dynamic linear model to detect dynamics in effect coefficients over time. We take our analysis of U.S. defense spending in two stages. First, using the dynamic estimation approach advocated here, we analyse U.S. spending based on similar data as True (2002), but for a longer time period, 1956-2010.[9] Second, we broaden the range of explanatory variables in light of the outcomes of the model 1 analysis and in light of the broader literature on U.S. defense spending. We hold constant across both model

specifications the response variable: the percentage of annual change in U.S. defense budget authority.[10] Thus, we collected similar data but for an expanded period, 1956-2010.

Table 1 provides an overview of the range and specification of explanatory variables used in models 1 and 2, respectively. As argued above, the dominant theories of public policy do not generally offer guidance about when relationships change, only that they may change over time. For instance, partisan alignments around the defense issue may change over time, reflecting new problem developments, a new understanding of the issue and/or new linkages of defense to other issues such as terrorism, the economy and employment. Similarly, the importance of public opinion on this issue relative to other drivers of policy making may change over time resulting in a dynamic impact of public opinion on defense spending. On the other hand, it could also be the case that such transformative change is less prevalent than claimed by theories of public policy. The previous time-series analyses on defense spending do not tell us much about this question and it is therefore warranted to initiate a closer inspection of this question utilizing the rich time series data on U.S. defense spending.

Summary statistics for all variables in both models can be found in appendix C. For both models, we summarize results by plotting estimated effect coefficients over time with confidence intervals. Since the model estimates coefficients and standard errors for each year (observation) in the data, coefficient tables are impractical. Results are most easily interpreted graphically. Both models use data standardized by subtracting variable means and dividing by one standard deviation.[11] As a result, both models also exclude constants. For assessments of both models' in-sample predictive accuracy, see appendix E.

Results Model 1

Dynamic coefficient estimates from our first model specification are plotted in figure 1. There is some initial evidence here that influences of public spending do not have constant effects over time. Periods of war and international tension are not consistently related to changes in defense spending, for example, with certain periods associated with more and

Table 1: Explanatory variables from previous research

Variable	Measure	Model 1[*]	Model 2[†]
Lag defense spending	Lagged real defense budget authority (True 2002)	✓	✓
International aid	U.S. foreign aid spending (constant USD) (True 2002)	✓	
War/tension	Indicator taking value of 1 in periods of war (i.e. Korean War, Vietnam War, Reagan buildup, first Gulf war, and 2001 onward) (True 2002).	✓	
Lag unemployment	National unemployment rate (True 2002)	✓	✓
Pres. election year	Indicator taking value of 1 in years in which an incumbent competes in presidential election (True 2002)	✓	✓
Lag Soviet spending	Lag Soviet defense spending assembled from U.S. intelligence reports (True 2002)	✓	
Lag Congressional ideology	Lagged polarization in the House of Representatives, i.e. mean DW-NOMINATE score on first dimension (Rosenthal & Poole 2015)		✓
Change in GDP	% change in GDP (Whitten & Williams 2011)		✓
Hostilities	Summed annual intensity of all U.S. militarized interstate disputes (Whitten & Williams 2011, Palmer, D'Orazio, Kenwick & Lane 2015)		✓
Lag public opinion	% of Americans saying foreign affairs is an important problem[‡] (Cusack 1992)		✓

Note: A more comprehensive list of explanatory variables from previous research can be found in appendix B

∗ This model builds on True (2002).

† This model builds on the outcome of the model 1 analysis as well as the broader literature on defense spending (see also table A.1 in appendix B).

‡ Public opinion data originally collected by Frank R. Baumgartner and Bryan D. Jones, with support of National Science Foundation grants SBR 9320922 and 0111611, distributed through Department of Government at the University of Texas at Austin. Neither NSF nor the original collectors of the data bear any responsibility for the analysis reported here.

certain periods with less spending. Furthermore, lagged real defense spending (in levels) is not consistently negatively associated with subsequent spending changes, which indicates a dynamic adjustment process (see De Boef & Keele 2008). Whereas a negative coefficient may be interpreted as reflecting negative feedback, with the defense budget regressing back to its "normal" level, a positive coefficient would signal positive feedback in the sense that high budgets would be followed by further increases (see also Wood & Doan 2003). Figure 1 does not show a full shift from negative to positive feedback, but the varying effects over time of the lagged spending measure may be interpreted as evidence in support of the

emphasis on changing feedback processes in the punctuated equilibrium theory (Baumgartner & Jones 2002).

The rest of the estimates from the model show little evidence of dynamic effects. International aid demonstrates a brief positive association with defense spending in the first decade of the 21st century. This largely reconfirms True's (2002) null finding regarding his hypothesis that international aid and military spending were determined in coordination. The period of our data also present in his analysis shows virtually no evidence in favor of this. Lagged unemployment shows statistically indistinguishable effects throughout, despite its mean estimated effect crossing the zero line during our period of analysis. The effects of both lagged Soviet defense spending and the election year indicator are basically constant and statistically indistinguishable from zero.

Since the combination of relatively short time series and numerous parameters being estimated places high demands on our limited amount of data, we may improve our estimates if we can eliminate unimportant information from the model. Thus, we now turn to our revised model specification.

Results Model 2

In our second model, plotted in figure 2, we see more suggestive evidence of dynamic effects over time. The results in this second model specification are estimated with greater certainty than the first and further sharpen some of the trends glimpsed in the first model. First, consistent with results in Figure 1, Figure 2 shows lagged defense spending is only sometimes negatively associated with percentage change in defense spending. Substantially, the dynamic estimates on lagged defense spending suggest that during the second half of the 1980s and through most of the 1990s the defense budget was regressing back to a "normal" level following a peak at the height of the Cold War. That is, the higher than average lagged spending levels in this period had a negative relationship to spending changes, and, indeed, the defense budget decreased annually from 1986 through 1998.
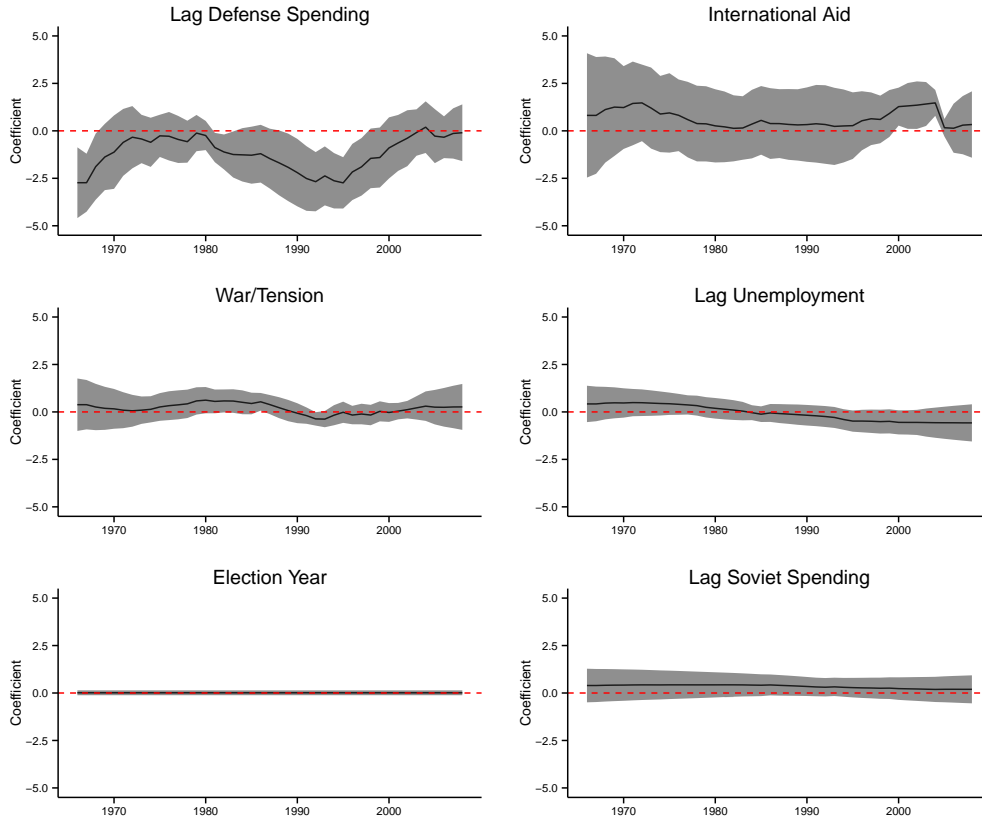
18

Figure 1: Estimated time-varying coefficients from model one

Note: Effect coefficients are plotted over time. The solid line indicates the point estimate for the coefficient each period, while gray areas are 95% confidence intervals. The dashed horizontal line indicates zero. Where the gray area overlaps the dashed line, estimates are not statistically significant by conventional standards.

Second, international hostilities display less dynamics than effects for True's war and tensions indicator in the previous model, yet it reaches negative statistical significance around the same time as the war/tension indicator — in the early 1990s. Although perhaps counter to intuition, the negative association between intensity of U.S. participation in foreign hostilities and defense spending changes is likely meaningful here. Hostilities varies annually and is, therefore, measured more finely than an indicator. However, the summary statistics show it does not vary widely — U.S. participation in militarized interstate disputes peaks several times during our period of analysis. Therefore, these estimates tell us that, despite an uptick in U.S. participation in militarized interstate disputes in the 1990s, the period of hostilities following the end of the Cold War coincided with decreasing military spending, in
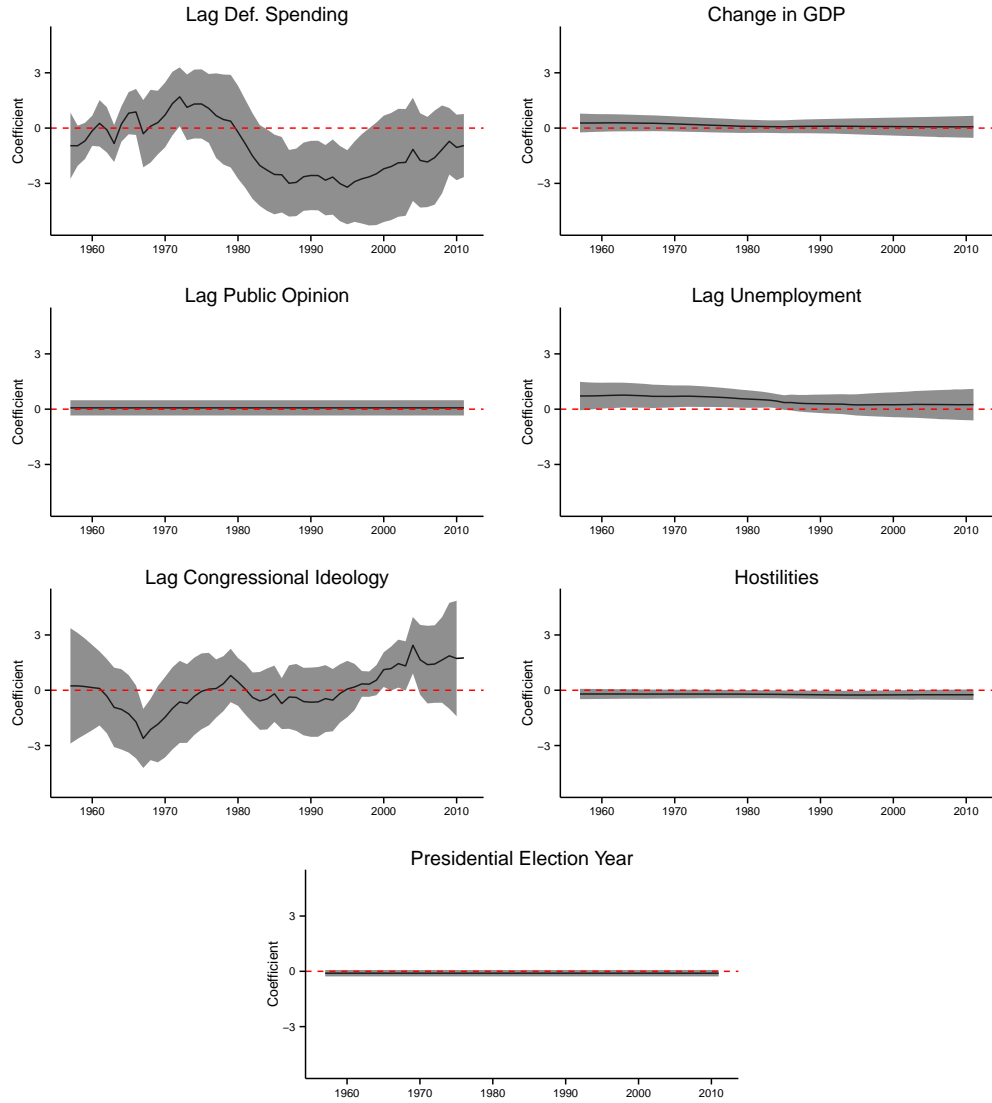
19

Figure 2: Estimated time-varying coefficients from model two

Note: Effect coefficients are plotted over time. The solid line indicates the point estimate for the coefficient each period, while gray areas are 95% confidence intervals. The dashed horizontal line indicates zero. Where the gray area overlaps the dashed line, estimates are not statistically significant by conventional standards.

contrast to the rest of the time series.

Third, the perhaps most noteworthy result in Figure 3 is that the effect of political preferences in the House of Representatives actually changes its direction over time. In contrast with classic conceptions of party positions on the defense issue, a more conservative House ideology had a negative association with changes in defense spending in the 1960s. The

estimate became indistinguishable from zero until the late 1990s, and then the association reversed so that more conservative House ideology was related to greater spending through the first few years of the new century. These results indicate that partisan politics intrude into defense spending changes only irregularly and in changing ways over time. Furthermore, the effects can be decisive. At its greatest positive point, in 2004, a one standard deviation shift in House ideology toward conservatism would have an estimated effect of increasing defense spending by around 20 percent. The actual shift in House ideology that year was around one fourth of a standard deviation toward the conservative side. At its most negative, in 1967, a shift toward conservatism in the House was associated with an even greater in magnitude *drop* in defense spending.

The timing of these effect estimates is noteworthy. The most negative estimate falls in 1967, following the major escalation of U.S. involvement in Vietnam after the Gulf of Tonkin Resolution, while the most positive estimate falls in 2004, as the depth of U.S. involvement required in Iraq was becoming clear the year after declaring victory. At these times of major foreign escalations, the partisanship of the House of Representatives is likely most important to presidents' ability to mobilize resources for increases in the defense budget. After all, the late 1960s and the early 2000s were both periods of increasing defense budgets to accommodate expanding conflicts and they were also periods in which the presidents could rely on co-partisan majorities in the House. Such partisan dynamics are in line with recent qualitative, comparative studies (e.g. Mortensen, Green-Pedersen, Breeman, Chaqus-Bonafont, Jennings, John, Palau & Timmermans 2011), but have not gained much scholarly attention in time-series analyses of public policy and public spending.

Fourth, the effect of the lagged national unemployment rate also merits attention. This effect estimate shows evidence of minor gradual changes in effect, going from a significant relationship to defense spending until the period of the Reagan build-up, after which it no longer shows a statistical relationship to changes in defense spending. Thus, our model indicates that the traditional expectation that increasing defense spending is a way to offset

unemployment finds support throughout most of our time series up until the last year of the Reagan buildup. However, after this point, we no longer find a statistically significant relationship. This finding is a strong candidate for a case in which politicians' core beliefs about policy have evolved. By the time of the defense buildup in the first part of the new century, several trends made defense spending a less useful tool for addressing unemployment: responding to international terrorism lent itself less to massive buildups of troops and bases, and sizable portions of defense spending were outsourced to private contracts. Thus, the logic underlying the former connection between unemployment and defense spending altered in the mid-1980s and seemingly has not returned.

Finally, public sentiment that defense is an important problem shows no association to defense spending changes, being consistently statistically insignificant and near zero. Likewise, presidential election years show no relationship to the outcome either.

DLM accuracy using simulated data

The model description demonstrates the DLM is capable of estimating coefficients that change with time, and our empirical examples show how these models can help us learn about policy making trends. Nevertheless. it is another matter to ask whether the dynamic estimates from a DLM are necessarily the "true" coefficients we would like to recover from the estimation. We examine this latter, more crucial, question by simulating a time series of 41 observations of data on which to test our model. The chosen number of observations reflects a the length of our actual time series from figure 1. We simulate a mix of six dynamic and constant true effect parameters and use them to calculate a simulated outcome variable that is a function of six randomly simulated explanatory variables and a normally distributed error term.

The design of the simulation reflects different types of time-varying relationships implied by policy theories. Whereas the framework of punctuated equilibrium theory claims that

changes can be sudden and quite dramatic, the advocacy coalition framework depicts more gradual changes, and a similar shape of change would be expected from the social learning perspective. In addition, it is important to validate that the DLM can in fact return constant effects where appropriate. Thus, two of the six factors are related to the outcome via constants (Beta 1 and Beta 5). One is constant except for a single large shift from a negative to a positive effect (Beta 6), which is chosen in order to examine whether the approach can pick up a major and sharp punctuation in the relationship between the explanatory variable and the outcome variable. One is related to the outcome via a coefficient that varies randomly according to a normal distribution over time but has no systematic pattern (Beta 4). Finally, two of these factors are related to the outcome via effect coefficients exhibiting distinct, gradual temporal patterns (Beta 2 and Beta 3).

The factors themselves, our simulated explanatory variables, are a matrix of six random vectors drawn from standard normal distributions. The outcome variable is calculated deterministically from the data and the true effects we wish to recover, then we add a realization of mean-zero normally distributed random noise to each observation. We exclude an intercept term from the model and, prior to running the model, we standardize the outcome by subtracting its mean and dividing by its standard deviation. We do not transform the explanatory variables, since they are random draws from standard normal distributions.

Figure 3 shows the results of running our DLM on this simulated data. The solid lines plot the true coefficients for our model and the dashed lines plot our estimates of them with confidence intervals in gray. The model is quite successful at tracking how coefficients evolve over time, in the case of the two dynamic coefficients (Beta 2 and Beta 3), and it filters out white noise to estimate Beta 4 as roughly constant. Beta 1 and Beta 5 are estimated accurately as constants, though the estimate for Beta is biased slightly downward. Finally, the model converges to a roughly correct estimate of Beta 6, though it estimates the last few observations somewhat too conservatively.

We note that repeating this test with all coefficients varying dynamically, or none varying
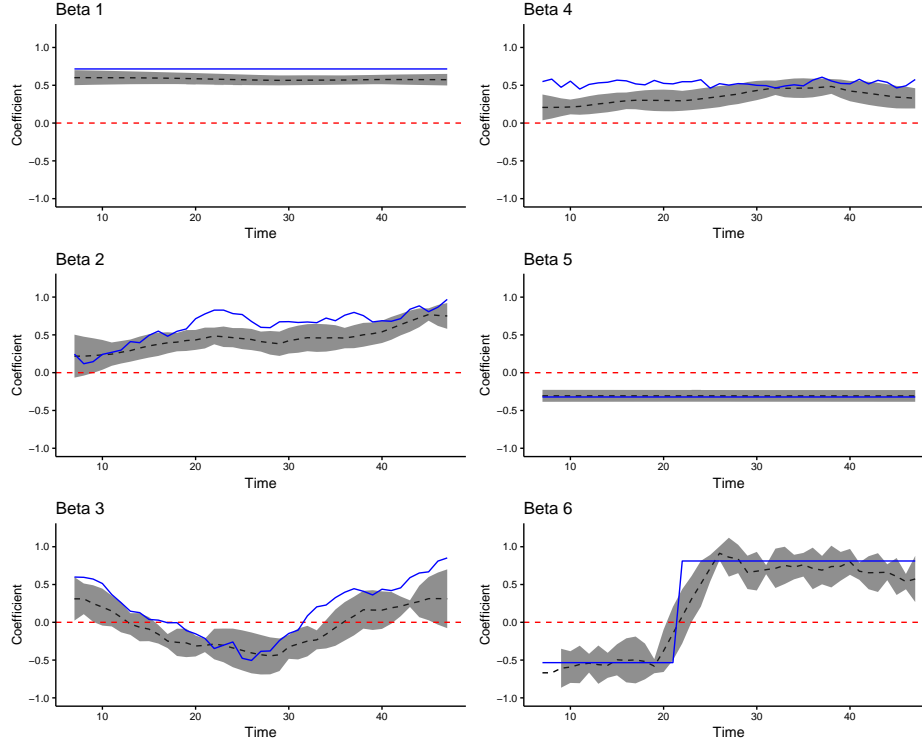
Figure 3: Estimated and true time-varying coefficients from simulated data

Note: The solid line is the known value of the coefficient. The dashed lines plot central predictions of dynamic effect coefficients, and the gray regions are 95% confidence intervals around those predictions.

dynamically, does not substantively change the accuracy. Maintaining the same patterns of dynamics and the same number of independent variables, we can also demonstrate that changing the length of the time series has relatively little impact on model performance. We refer the reader to appendix G for graphs of these models in which we vary the length of the time series from a low of 30 observations to a high of 100 observations. Results are virtually identical to those reported here.

These are important evidence demonstrating that the DLM modeling approach is sensitive to various types of transformative changes, both the gradual movement predicted by the social learning and advocacy coalition framework, as well as the more punctuated changes of main interest to the punctuated equilibrium theory. It should be noted, however, that these findings do not amount to conclusive evidence that the DLM will recover the true coefficients under all circumstances, only that the DLM can recover true estimates under

conditions similar to our own empirical tests. We encourage the reader to perform simulations emulating the patterns of variation in their own data when applying the DLM. To aid in this, we provide example code in appendix H for creating simulations like our own using the R statistical software environment and the "dlm" package (Petris 2010).

Extending DLMs to panel data

Sometimes policy scholars utilize panel data, for instance several countries observed over time, to evaluate dynamic policy theories. Extending DLMs using panel data can be accomplished in several ways. Two common choices are to model the panel time series in the DLM framework either as seemingly unrelated time series equations (SUTSE) or by using a hierarchical model with levels accommodating panel variation. In the former approach, the model structure is virtually the same as above except that errors in both the measurement and state equations are assumed correlated across panel units within time periods (Petris, Petrone & Campagnoli 2009, pp. 132–134). This amounts to a partial relaxation of the assumption that all errors are independent. Errors in a SUTSE DLM model take a block diagonal form, in which errors within a period across units are correlated and all other errors are independent. It can, however, become computationally demanding to estimate the large covariance matrices that result when the number of panel units is large.

In such cases, an alternative is to estimate a dynamic hierarchical model. In this framework, the state equation has (at least) two levels. The dynamic coefficient estimates (in measurement equation) are estimated as realizations from an underlying state (first state level), and the state varies as a Markov chain over time (second state level) (Petris, Petrone & Campagnoli 2009, 134–136). This introduces dependence across panel units, which also varies dynamically, without requiring estimation of large covariance matrices. In such a

framework, with panel units denoted by j, a basic model would be:

$$y_{jt} = \boldsymbol{X_{jt}}\boldsymbol{\beta_{jt}} + v_{jt}, \qquad\qquad v_t \sim i.i.d.(0, \sigma^2_{jt}),$$

$$\boldsymbol{\beta_{jt}} = \boldsymbol{\lambda_t} + \boldsymbol{\epsilon_{jt}}, \qquad\qquad \boldsymbol{\epsilon_{jt}} \sim i.i.d.(0, \Sigma_t),$$

$$\boldsymbol{\lambda_t} = \boldsymbol{\lambda_{t-1}} + \boldsymbol{w_t}, \qquad\qquad \boldsymbol{w_t} \sim i.i.d.(0, Q)$$

It remains as Commandeur, Koopman & Ooms (2011) noted in 2011, that, at the time of writing, not all software packages for estimating DLMs support every form of DLM. Readers who find their preferred software does not include commands for easily estimating dynamic hierarchical or seemingly unrelated regressions can find this functionality in the "dlm" package for R. Similarly, readers who wish to estimate dynamic non-linear regressions will find that, at the time of writing, "dlm" does not provide commands for this while MATLAB, RATS, or the "KFAS" R package do provide such tools.

Conclusion

Theories of public policy imply that the causes of public policy may not be consistent across time. These time-varying relationships represent a major challenge to standard regression techniques, since standard techniques assume that a change in the value of an explanatory variable, x, will produce a corresponding change in the value of a dependent variable, y, of the same magnitude and direction at all times. This lack of alignment between our policy theories, which are explicitly dynamic, and our methods of empirical analysis, which are largely static, has been forcefully pointed out by leading public policy scholars. These voices have advocated for increased use of alternative methods such as process tracing and stochastic process analyses (Jones & Baumgartner 2005, Hall 2006).

While we applaud this expansion of methods within the field of public policy research, this paper has advocated a third alternative: better aligning statistical time series analysis

of public policy with the ontology of the major theories in the field. More particularly, we have advocated the DLM – a flexible approach to time-series analysis that can uncover both time-varying and constant relationships. Compared to alternative dynamic statistical methods utilized in public policy research, DLMs require fewer assumptions from the analyst, allowing the data to speak for itself to the greatest extent regarding the timing, direction, and magnitude of effect changes. This is a major advantage over alternative methods, given the lack of specificity regarding the timing and particular conditions of change in public policy theories.

Another advantage is the fact that DLMs utilize all the information in the time-series data to model policy outcomes in a way that closely maps onto the data-generating process envisioned by our theories of public policy. Given the multiple causes of public policies, a further advantage of the modeling approach advocated in this paper is the ability to estimate multivariate models. Furthermore, because DLMs provide information about both stable and time-varying coefficients, they can be used to investigate the general expectation derived from all of the major policy theories that relationships are time-varying. This is clearly illustrated in our example analysis of U.S. defense spending, where a mix of constant, dynamic and non-systematic effects are present.

Applying the DLM to policy analysis opens up a range of new avenues for public policy research. First, whereas much quantitative policy research has been focused on theorizing and estimating stable relationships, applying DLMs as a tool invites a higher order level of theorizing and modeling aimed at explaining when and why the relationship between x and y may change. Inspired by Lieberman (2002), this latter type of change can be denoted *transformative change.* Given the flexibility of the DLM, it is also a tool that can aid in theory development. As we identify transformative changes in relationships between policy inputs and policy outputs, these findings – and the further puzzles they will present – will direct and inform refinements in our theoretical explanations of the policy process.

A related strand of research our work points to is the usefulness of applying DLMs on a

larger scale to identify variables, or sets of variables, most prone to exhibiting time-varying relationships to policy outcomes. As DLMs are applied to more policy areas and/or more countries, we hope this will reveal which types of variables – for instance, political, economic, or institutional – exhibit more dynamic effects. Though it may remain futile to attempt to make point predictions about the exact form and timing of transformative changes, it may eventually be possible – and would certainly be useful for future theorizing – to identify clusters of more or less stable causes of public policy. Developing a systematic understanding of when dynamic relationships are observed will help us refine our theories about how they arise in the first place.

A third extension would be to examine time-series data that are closer to the actual policymaking process, such as agenda setting data covering congressional hearings, presidential speeches, law-making activities, and other outcomes that are intermediate to the production of actual policy outputs. The fact that such data is now available for the U.S., as well as a number of other countries,[12] means that we can utilize DLMs to begin comparing transformative dynamics across different stages of the policy making process and across political systems.

As our examples have illustrated, the DLM can be a powerful tool for studying dynamic relationships. We believe that applying DLMs more widely offers significant practical and theoretical advantages to students of policy change in the work of probing and expanding our leading explanations of stability and change in public policy.

Notes

[1]See also Weible, Sabatier & McQueen (2009) and Weible, Sabatier, Jenkins-Smith, Nohrstedt, Henry & deLeon (2011).

[2]In a broader perspective, dynamic causal forces have also been a major theme in comparative politics, exemplified by research on historical institutionalism (Steinmo, Thelen & Longstreth 1992, Lieberman 2002, Pierson 2000) as well as reflected in more particular discussions about the changing effect of political parties on welfare policies (Pierson 1996, Huber & Stephens 2001, Kwon & Pontusson 2010).

[3]These approaches are algebraically equivalent approaches to estimating dynamic coefficients in linear models (Montana, Triantafyllopoulos & Tsagaris 2009)

[4]There is also ongoing debate regarding the econometric properties of DCC/MV-GARCH estimators (Caporin & McAleer 2013).

[5]See appendices D and E.

[6]An example of our main results when applying filtering only can be seen in appendix I, for reference.

[7]See also, (Petris, Petrone & Campagnoli 2009, pp. 34, 115).

[8]See Correa & Kim (1992) for a similar review containing some older material we have omitted.

[9]Lagged Soviet defense spending is only available between 1966 and 2007 (Russian defense spending is substituted following the Soviet Union's collapse). We collected data from the original U.S. intelligence estimates since we were unable to procure True's own data. See appendix F for a replication on True's exact time period and measurement choices.

[10]Budget authority data were originally collected by Frank R. Baumgartner and Bryan D. Jones, with the support of National Science Foundation grant numbers SBR 9320922 and 0111611, and were distributed through the Department of Government at the University of Texas at Austin. Neither NSF nor the original collectors of the data bear any responsibility for the analysis reported here.

[11]See appendix C for plots of standardized variables.

[12]See `www.comparativeagendas.net`.

References

Baltagi, Badi. 2011. *Econometrics.* Springer Texts in Business and Economics.

Baumgartner, Frank R. & Bryan D. Jones. 1993. *Agendas and Instability in American Politics.* Chicago, IL: University of Chicago Press.

Baumgartner, Frank R. & Bryan D. Jones. 2015. *The Politics of Information: Problem Definition and the Course of Public Policy in America.* University Of Chicago Press.

Baumgartner, Frank R. & Bryan D. Jones, eds. 2002. *Policy Dynamics.* University Of Chicago Press.

Baumgartner, Frank R, Bryan D Jones & Peter B Mortensen. 2014. *Theories of the Policy Process.* Westview Press chapter Punctuated equilibrium theory: Explaining stability and change in public policymaking, pp. 59–103.

Beck, Nathaniel. 1983. "Time-Varying Parameter Regression Models." *American Journal of Political Science* 27(3):557–600.

Blyth, Mark. 2013. "Paradigms and Paradox: The Politics of Economic Ideas in Two Moments of Crisis." *Governance* 26(2):197–215.

Brandt, Patrick T., John T. Williams, Benjamin O. Fordham & Brain Pollins. 2000. "Dynamic Modeling for Persistent Event-Count Time Series." *American Journal of Political Science* 44(4):823–843.

Cairney, Paul & Tanya Heikkila. 2014. *Theories of the Policy Process.* Third edition ed. Chicago: Westview Press chapter A Comparison of Theories of the Policy Process.

Campbell, John L. 2002. "Ideas, Politics, and Public Policy." *Annual Review of Sociology* 28(1):21–38.

Caporin, Massimiliano & Michael McAleer. 2013. "Ten Things You Should Know about the Dynamic Conditional Correlation Representation." *Econometrics* 1(1):115–126.

Commandeur, Jacques J. F. & Siem Jan Koopman. 2007. *An Introduction to State Space Time Series Analysis.* Oxford University Press.

Commandeur, Jacques J. F., Siem Jan Koopman & Marius Ooms. 2011. "Statistical Software for State Space Methods." *Journal of Statistical Software* 41(1).

Correa, Hector & Ji-Won Kim. 1992. "A Causal Analysis of the Defense Expenditures of the USA and the USSR." *Journal of Peace Research* 29(2):161–174.

Cusack, Thomas R. 1992. *The Political Economy of Military Spending in the United States.* London: chapter On the Domestic Political-Economic Sources of American Military Spending.

Cusack, Thomas R. & Michael Don Ward. 1981. "Military Spending in the United States, Soviet Union, and the People's Republic of China." *The Journal of Conflict Resolution* 25(3):429–469.

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

Engle, Robert. 2002. "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models." *Journal of Business & Economic Statistics* 20(3):339–350.
URL: *http://www.jstor.org/stable/1392121*

Enns, Peter K. & Gregory E. McAvoy. 2011. "The Role of Partisanship in Aggregate Opinion." *Political Behavior* 34(4):627–651.

Fariss, Christopher J. 2014. "Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability." *American Political Science Review* 108:297–318.

Finseraas, Henning & Kåre Vernby. 2011. "What parties are and what parties do: partisanship and welfare state reform in an era of austerity." *Socio-Economic Review* 9(4):613–638.

Gerber, Alan & Donald P. Green. 1998. "Rational Learning and Partisan Attitudes." *American Journal of Political Science* 42(3):794–818.

Green, Donald P., Alan S. Gerber & Suzanna L. de Boef. 1999. "Tracking Opinion Over Time: A Method for Reducing Sampling Error." *The Public Opinion Quarterly* 63(2):178–192.

Griffin, L. J., M. Wallace & J. Devine. 1982. "The political economy of military spending: evidence from the United States." *Cambridge Journal of Economics* 6(1):1–14.

Hall, Peter A. 1993. "Policy paradigms, social learning, and the state: the case of economic policymaking in Britain." *Comparative Politics* (25):275–296.

Hall, Peter A. 2003. *Comparative Historical Analysis in the Social Sciences.* Cambridge, UK and New York: Cambridge University Press chapter Aligning Ontology and Methodology in Comparative Research, p. Ch. 11.

Hall, Peter A. 2006. "Systematic process analysis: when and how to use it." *European Management Review* 3(1):24–31.

Hartley, Thomas & Bruce Russett. 1992. "Public Opinion and the Common Defense: Who Governs Military Spending in the United States?" *The American Political Science Review* 86(4):905–915.

Higgs, Robert & Anthony Kilduff. 1993. "Public Opinion: A Powerful Predictor of U.S. Defense Spending." *Defence Economics* (4):227–238.

Hogan, John & Michael Howlett, eds. 2015. *Policy Paradigms in theory and Practice. Discourses, Ideas and Anomalies in Public Policy Dynamics.* Palgrave Macmillan.

Huber, Evelyne & John D. Stephens. 2001. *The New Politics of the Welfare State.* Oxford University Press chapter Welfare State and Employment Regimes in the Era of Retrenchment, pp. 107–45.

Jenkins-Smith, Hank C., Daniel Nohrstedt, Christopher M. Weible & Paul A. Sabatier. 2014. *Theories of the Policy Process.* 3rd ed. Westview Press chapter The Advocacy Coalition Framework: Foundations, Evolution, and Ongoing Research, pp. 183–223.

Jones, Bryan D. 1994. *Reconceiving Decision-Making in Democratic Politics: Attention, Choice, and Public Policy.* University Of Chicago Press.

Jones, Bryan D. & Frank R. Baumgartner. 2005. *The politics of attention. How Government Prioritizes Problems.* Chicago: University of Chicago Press.

Kalaba, Robert & Leigh Tesfatsion. 1988. "The Flexible Least Squares Approach to Time-Varying Linear Regression." *Journal of Economic Dynamics and Control* 12:43–48.

Kamlet, Mark S. & David C. Mowery. 1987. "Influences on Executive and Congressional Budgetary Priorities, 1955-1981." *The American Political Science Review* 81(1):155–178.

Kiewiet, D. Roderick & Mathew D. McCubbins. 1991. *The Logic of Delegation: Congressional Parties and the Appropriations Process.* University Of Chicago Press.

Kim, Chang-Jin & Charles R. Nelson. 2000. *State-Space Models with Regime Switching.* Massachusetts Institute of Technology.

Kwon, Hyeok Yong & Jonas Pontusson. 2010. "Globalization, labour power and partisan politics revisited." *Socio-Economic Review* 8(2):251–281.

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

Lieberman, Robert C. 2002. "Ideas, Institutions, and Political Order: Explaining Political Change." *The American Political Science Review* 96(4):697–712.

Majeski, Stephen J. 1983. "Mathematical Models of the U.S. Military Expenditure Decision-making Process." *American Journal of Political Science* 27(3):485–514.

Majeski, Stephen J. 1992. *The Political Economy of Military Spending in the United States.* London: Routledge chapter Defense Budgeting, Fiscal Policy, and Economic Performance.

Mcavoy, Gregory E. 2006. "Stability and Change: The Time Varying Impact of Economic and Foreign Policy Evaluations on Presidential Approval." *Political Research Quarterly* 59(1):71–83.

Mergner, Sascha. 2009. *Applications of State Space Models in Finance.* Universittsverlag Gttingen.

Mitchell, Sara McLaughlin, Scott Gates & Håvard Hegre. 1999. "Evolution in Democracy-War Dynamics." *Journal of Conflict Resolution* 43(6):771–792.

Montana, Giovanni, Kostas Triantafyllopoulos & Theodoros Tsagaris. 2009. "Flexible least squares for temporal data mining and statistical arbitrage." *Expert Systems with Applications* 36(2, Part 2):2819–2830.

Mortensen, Peter Bjerre, Christoffer Green-Pedersen, Gerard Breeman, Laura Chaqus-Bonafont, Will Jennings, Peter John, Anna M. Palau & Arco Timmermans. 2011. "Comparing Government Agendas: Executive Speeches in the Netherlands, United Kingdom, and Denmark." *Comparative Political Studies* 44(8):973–1000.

Nincic, Miroslav & Thomas R. Cusack. 1979. "The Political Economy of US Military Spending." *Journal of Peace Research* 16(2):101–115.

Ostrom, Charles W., Jr. 1977. "Evaluating Alternative Foreign Policy Decision-Making Models: An Empirical Test between an Arms Race Model and an Organizational Politics Model." *The Journal of Conflict Resolution* 21(2):235–266.

Ostrom, Charles W., Jr. 1978. "A Reactive Linkage Model of the U.S. Defense Expenditure Policymaking Process." *The American Political Science Review* 72(3):941–957.

Ostrom, Charles W., Jr. & Robin F. Marra. 1986. "U.S. Defense Spending and the Soviet Estimate." *The American Political Science Review* 80(3):819–842.

Palmer, Glenn, Vito D'Orazio, Michael Kenwick & Matthew Lane. 2015. "The MID4 dataset, 20022010: Procedures, coding rules and description." *Conflict Management and Peace Science* 32(2):222–242.

Petris, Giovanni. 2010. "An R Package for Dynamic Linear Models." *Journal of Statistical Software* 36(12):1–16.

Petris, Giovanni & Sonia Petrone. 2011. "State Space Models in R." *Journal of Statistical Software* 41(4).

Petris, Giovanni, Sonia Petrone & Patrizia Campagnoli. 2009. *Dynamic Linear Models with R.* Springer.

Pierson, Paul. 1996. "The New Politics of the Welfare State." *World Politics* 48:143–179.

Pierson, Paul. 2000. "Path Dependence, Increasing Returns, and the Study of Politics." *The American Political Science Review* 94(2):251–267.

Princen, Sebastiaan & Paul 't Hart. 2014. "Putting Policy Paradigms in Their Place." *Journal of European Public Policy* 21(3):470–474.

Rosenthal, Howard & Keith Poole. 2015. "House Polarization 1st to 113th Congresses.".
    URL: *ftp://voteview.com/house_polarization46_113.txt*

Sabatier, Paul. 1987. "Knowledge, policy-oriented learning, and policy change." *Knowledge* 8:649–692.

Sabatier, Paul. 1988. "An advocacy coalition framework of policy change and the role of policy-oriented learning therein." *Policy Sciences* 21:129–168.

Sabatier, Paul A. 1998. "The advocacy coalition framework: revisions and relevance for Europe." *Journal of European Public Policy* 5(1):98–130.

Sabatier, Paul & Hank Jenkins-Smith, eds. 1993. *Policy Change and Learning: An Advocacy Coalition Approach.* Boulder, CO: Westview Press.

Shumway, Robert H. & David S. Stoffer. 2010. *Time Series Analysis and Its Applications: With R Examples.* Springer.

Steinmo, Sven, Kathleen Thelen & Frank Longstreth, eds. 1992. *Structuring Politics: Historical Institutionalism in Comparative Analysis.* Cambridge University Press.

Su, Tsai-Tsu, Mark S. Kamlet & David C. Mowery. 1993. "Modeling U.S. Budgetary and Fiscal Policy Outcomes: A Disaggregated, Systemwide Perspective." *American Journal of Political Science* 37(1):213–245.

True, James L. 2002. *Policy Dynamics.* Chicago: University of Chicago Press chapter The Changing Focus of National Security Policy, pp. 155–183.

Tusell, Fernando. 2011. "Kalman Filtering in R." *Journal of Statistical Software* 39(2).

Weible, Christopher M., Paul A. Sabatier, Hank C. Jenkins-Smith, Daniel Nohrstedt, Adam Douglas Henry & Peter deLeon. 2011. "A Quarter Century of the Advocacy Coalition Framework: An Introduction to the Special Issue." *The Policy Studies Journal* 39(3):349–360.

Weible, Christopher M., Paul A. Sabatier & Kelly McQueen. 2009. "Themes and Variations: Taking Stock of the Advocacy Coalition Framework." *The Policy Studies Journal* 37(1):121–140.

Whitten, Guy D. & 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):117–134.

Wlezien, Christopher. 1996. "Dynamics of Representation: The Case of US Spending on Defence." *British Journal of Political Science* 26(1):81–103.

Wood, B. Dan. 2000. "Weak Theories and Parameter Instability: Using Flexible Least Squares to Take Time Varying Relationships Seriously." *American Journal of Political Science* 44(3):603–618.

Wood, B. Dan & Alesha Doan. 2003. "The Politics of Problem Definition: Applying and Testing Threshold Models." *American Political Science Review* 47(4):640–653.

Zuk, Gary & Nancy R. Woodbury. 1986. "U.S. Defense Spending, Electoral Cycles, and Soviet-American Relations." *The Journal of Conflict Resolution* 30(3):445–468.

Supplemental Information (Not for Print Publication)

A   Appendix: Model details

The model exposition we adopt is that outlined in Shumway & Stoffer (2010). Their work provides a general version of the specific approach we utilize. The estimation of the time-varying covariates is recursive, starting at time one and progressing to time t. This entire recursive estimation is repeated with each of the maximum likelihood estimates of the additional parameters of the model, beginning with their initially chosen starting values and ending with the stable final estimates. Recall that our model is:

$$
\begin{aligned}
y_t &= \boldsymbol{X_t}\boldsymbol{\beta_t} + v_t \\
v_t &\sim i.i.d.(0, \sigma_v^2)
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\beta_t} &= \boldsymbol{\beta_{t-1}} + \boldsymbol{w_t} \\
\boldsymbol{w_t} &\sim i.i.d.(0, Q)
\end{aligned}
$$

The scalar y_t is the outcome in time t. Design vector $\boldsymbol{X_t}$ is a $1 \times k$ vector of explanatory variables measured at time t. Coefficient vector $\boldsymbol{\beta_t}$ is a $k \times 1$ vector of estimates at time t. The scalar v_t and vector $\boldsymbol{w_t}$ $(k \times 1)$ are mean-zero normally distributed noise. The parameters σ_v^2 and Q $(k \times k)$ are their respective variances. Based on these assumptions, estimation proceeds as follows:

1. Select initial values for the parameters: $\boldsymbol{\beta_0}$, Q, σ_v^2, and the covariance matrix of innovations (or prediction errors) Σ_0.

2. Run the Kalman filter to obtain values for the innovations (prediction errors) from the model, $\boldsymbol{v}$, and their covariance.

3. Use the estimates obtained from the Kalman filter to estimate $\boldsymbol{\beta_0}$, Q, σ_v^2, and Σ_0 using maximum likelihood.

4. Repeat step 2 using the estimates from step 3 in place of the starting values selected in step 1.

5. Repeat step 4 until the estimates of $\boldsymbol{\beta_0}$, Q, σ_v^2, and Σ_0 or the likelihood stabilizes.

The recursion for the Kalman filter proceeds as follows. At time 0, the following two steps are unique and take the place of steps one and two in the next list:

1. Calculate an expectation of $\boldsymbol{\beta_1}$, conditional on $\boldsymbol{\beta_0}$. We assume that $\boldsymbol{\beta}$ follows a random walk, therefore our expectation of its next period value is always its current estimate: $\boldsymbol{\beta_{t|t-1}} = \boldsymbol{\beta_{t-1|t-1}}$. Likewise: $\boldsymbol{\beta_{1|0}} = \boldsymbol{\beta_0}$

2. Calculate an expectation of the covariance of innovations to $\boldsymbol{\beta_1}$, conditional on Σ_0 and Q. We refer to this as P: $P_{1|0} = \Sigma_0 + Q$.

From time 1 through time t, the following steps are taken:

1. $\boldsymbol{\beta_{t|t-1}} = \boldsymbol{\beta_{t-1|t-1}}$

2. $P_{t|t-1} = P_{t-1|t-1} + Q$

3. Calculate the predicted value of y conditional on expectations from time $t-1$:
$y_{t|t-1} = \boldsymbol{X_t}\boldsymbol{\beta_{t|t-1}}$

4. Calculate the prediction error in time t: $\eta_{t|t-1} = y_t - y_{t|t-1}$

5. Calculate the Kalman gain for period t, i.e. the proportion of uncertainty in each parameter in $\boldsymbol{\beta_t}$ attributable to uncertainty regarding the parameter relative to the full uncertainty in the model:

$$K_t = \frac{P_{t|t-1}X_t'}{X_t P_{t|t-1}X_t' + \sigma_v^2}$$

6. Update estimate of effect coefficients, $\boldsymbol{\beta_{t|t}}$, based on prediction error and Kalman gain:

$$\boldsymbol{\beta_{t|t}} = \boldsymbol{\beta_{t|t-1}} + K_t \eta_{t|t-1}$$

The value of $\boldsymbol{\beta_{t|t}}$ is our estimate of $\boldsymbol{\beta_t}$.

7. Update expectation of the covariance of parameters, $P_{t|t}$, based on X_t and σ_v^2:

$$P_{t|t} = P_{t|t-1} - \frac{P_{t|t-1}X_t'}{X_t P_{t|t-1}X_t' + \sigma_v^2} X_t P_{t|t-1}$$

This can also be expressed as:

$$P_{t|t} = [I - K_t X_t] P_{t|t-1}$$

B Appendix: Explanatory variables from previous defense spending studies

Table A.1: Explanatory variables from previous work

Concept	Measure/data source	Citations
Inertia or fatigue/burden of defense spending	lagged real congressional defense appropriation; (lagged) defense spending as % of GDP; current military budget request; current presidential budget request; budget authority for defense; defense spending as % of total; lagged % change in defense spending; % change in ratio of defense price deflator to GDP deflator	Ostrom 1977, Cusack & Ward 1981, Griffin, Wallace & Devine 1982, Majeski 1983, Kamlet & Mowery 1987, Correa & Kim 1992, Cusack 1992, Hartley & Russett 1992, Majeski 1992, Higgs & Kilduff 1993, Su, Kamlet & Mowery 1993, Wlezien 1996, Whitten & Williams 2011
Labor interests	(lagged) national (union/non-union) unemployment rate; lagged predicted unemployment; avg. duration of unemployment in weeks; % of civilian labor force employed 15 weeks or longer; change in unemployment by quarter; an index of corporatism; % of work days lost to strikes; military conscription; % change in unionization rate	Griffin, Wallace & Devine 1982, Ostrom & Marra 1986, Kamlet & Mowery 1987, Kiewiet & McCubbins 1991, Kiewiet & McCubbins 1991, Correa & Kim 1992, Cusack 1992, Majeski 1992, Su, Kamlet & Mowery 1993, True 2002
Election cycles	indicator for year preceding or coinciding with on-year elections; indicator for year n following presidential election; indicator for presidential election year with incumbent competing, years to next presidential election	Nincic & Cusack 1979, Cusack & Ward 1981, Griffin, Wallace & Devine 1982, Zuk & Woodbury 1986, Kamlet & Mowery 1987, Cusack 1992, True 2002, Whitten & Williams 2011
Arms race/Soviet threat	lagged or current real Soviet defense spending; lagged change in Soviet defense spending; lagged (change in) Soviet minus U.S. military spending; lagged average U.S.-Soviet conflict score from Conflict and Peace Data Bank; ratio of Soviet to U.S. strategic warheads; Correlates of War composite indicator of national capability	Ostrom 1977, Cusack & Ward 1981, Griffin, Wallace & Devine 1982, Majeski 1983, Ostrom & Marra 1986, Kamlet & Mowery 1987, Correa & Kim 1992, Cusack 1992, Hartley & Russett 1992, Su, Kamlet & Mowery 1993, True 2002, Whitten & Williams 2011
War/tension/hostility	count of years since start of ongoing war; Correlates of War tension score; "war commitment" index decreasing with years of involvement; lagged (change in) U.S. service members killed in action; estimated (change in) U.S. war costs; estimated Soviet war costs; other NATO defense spending; other Warsaw Pact defense spending; sum of annual Correlates of War militarized interstate dispute hostility scores	Nincic & Cusack 1979, Cusack & Ward 1981, Griffin, Wallace & Devine 1982, Ostrom & Marra 1986, Zuk & Woodbury 1986, Kamlet & Mowery 1987, Correa & Kim 1992, Cusack 1992, Su, Kamlet & Mowery 1993, Whitten & Williams 2011

Public opinion	indicator of negative opinion on defense spending; % support for president; % saying economy is most important problem; inflation minus unemployment raised to the power of public opinion on the economy; % saying foreign affairs is most important problem; % saying Vietnam is most important problem; ratio of Soviet to U.S. defense spending raised to the power of the diff. in public opinion on foreign affairs and Vietnam; change in % saying U.S. spends too little on defense; (lagged) net support for defense spending; avg. preference for the same spending level or with no opinion; % supporting higher defense spending; % supporting higher low-income entitlements; (change in) lagged net dislike of Soviet Union	Ostrom & Marra 1986, Correa & Kim 1992, Cusack 1992, Hartley & Russett 1992, Higgs & Kilduff 1993, Su, Kamlet & Mowery 1993, Wlezien 1996
International aid	foreign aid expenditures	True 2002
Government ideology	party of the president; %/number of (Northern) Democrats in House (Senate); lagged % change in Democratic House seats	Griffin, Wallace & Devine 1982, Kamlet & Mowery 1987, Cusack 1992, Wlezien 1996
Economy	change in nat. consumption and investment spending; (lagged) inflation (by category of veteran benefits and by category of entitlements); indicator for recession year; predicted full employment GDP minus actual GDP (GDP gap); GDP gap, manufacturing only; U.S. (Soviet) real (nominal) GDP (per capita); % change GDP; % change in monopoly capital sector profits; inflation by quarter; GDP as % of total OECD GDP; money supply in constant (current) dollars; GDP deflator; % change in share of manufacturing assets held by largest 200 corporations; % change in poverty rate	Nincic & Cusack 1979, Griffin, Wallace & Devine 1982, Kamlet & Mowery 1987, Correa & Kim 1992, Cusack 1992, Majeski 1992, Su, Kamlet & Mowery 1993, Whitten & Williams 2011
General federal spending	civilian federal outlays as % of GDP; federal revenue as % of GDP; projected (actual) revenue; projected (actual) (change in) federal deficit; outlays (un)controllable by Congress (president); revenue in pres. budget proposal; revenue in Congress budget; % change in ratio of non-defense price deflator to GDP deflator; fiscal year indicators	Griffin, Wallace & Devine 1982, Majeski 1983, Ostrom & Marra 1986, Kamlet & Mowery 1987, Hartley & Russett 1992, Majeski 1992, Su, Kamlet & Mowery 1993

Note: Though we build on their analysis, Whitten & Williams (2011) do not consider U.S. defense spending. Therefore, we have included only their variables we consider meaningful in the U.S. context.

C Appendix: Summary statistics

	Min	Max	Mean	Std Dev	Obs
Year	1966.00	2007.00	1986.50	12.27	42.00
% Change in defense spending	-8.90	24.74	2.27	7.57	42.00
Lag Defense Spending	52075.00	556290.00	228542.57	131887.93	42.00
International Aid	2479.00	30513.00	7085.52	4853.28	42.00
War/Tension	0.00	1.00	0.48	0.51	42.00
Lag Unemployment	3.49	9.71	5.89	1.50	42.00
Presidential Election Year	0.00	1.00	0.17	0.38	42.00
Lag Soviet Spending	24300.00	317900.00	129718.76	99864.17	42.00

Table A.2: Summary Statistics for Model 1

	Min	Max	Mean	Std Dev	Obs
Year	1957.00	2010.00	1983.50	15.73	54.00
% Change in defense spending	-8.90	24.74	2.44	7.22	54.00
Lag Defense Spending	34983.00	697763.00	222615.26	174427.24	54.00
Change in GDP	-0.02	0.13	0.07	0.03	54.00
Lag Public Opinion	0.00	0.47	0.16	0.14	54.00
Lag Unemployment	3.49	9.71	5.86	1.46	54.00
Lag Congressional Ideology	-0.13	0.15	-0.02	0.08	54.00
Hostilities	5.00	39.00	17.74	6.54	54.00
Presidential Election Year	0.00	1.00	0.15	0.36	54.00

Table A.3: Summary Statistics for Model 2
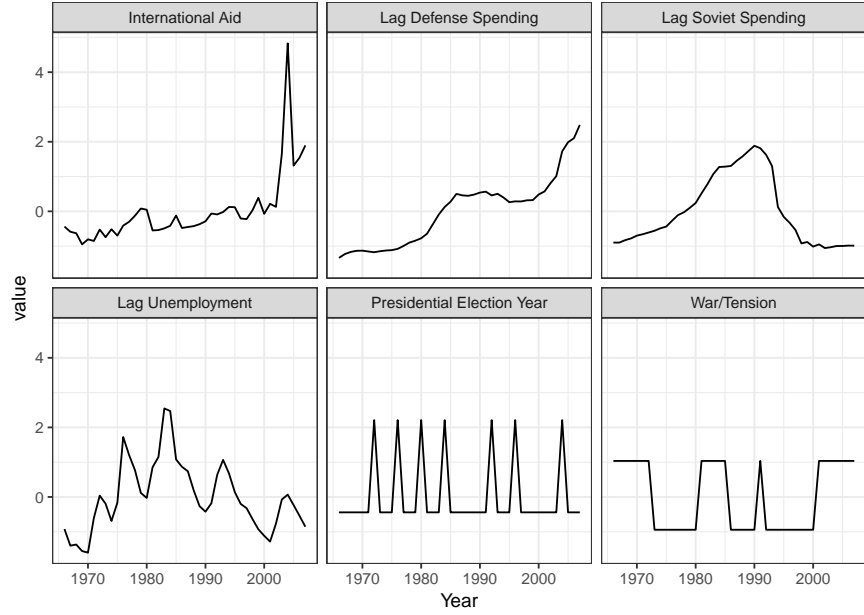
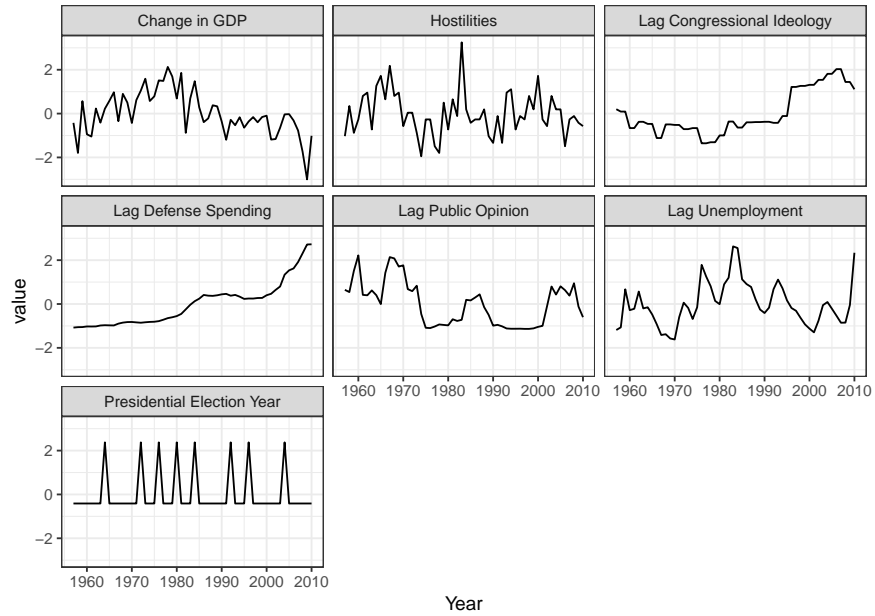Figure A.1: Standardized variables from model 1

Figure A.2: Standardized variables from model 2

A8

D   Appendix: Diagnostic tests on main model

In this section, we assess the primary assumptions underlying the DLM in our empirical application. These are: independence, homoskedasticity, and normality of the (standardized) residuals. Significance tests in the context of DLMs are typically much more reliable than in standard linear regression models, precisely because the residuals are generally closer to satisfying the assumption that they are independent random values (Commandeur & Koopman 2007, p. 158). Nevertheless, this feature of the DLM should be substantiated with appropriate diagnostic tests to ensure that the residuals are appropriately behaved. We report here appropriate diagnostic tests for our second model of U.S. defense spending, depicted in figure 2. The standardized residuals from this, our main model, can be seen in figure A.3 below.
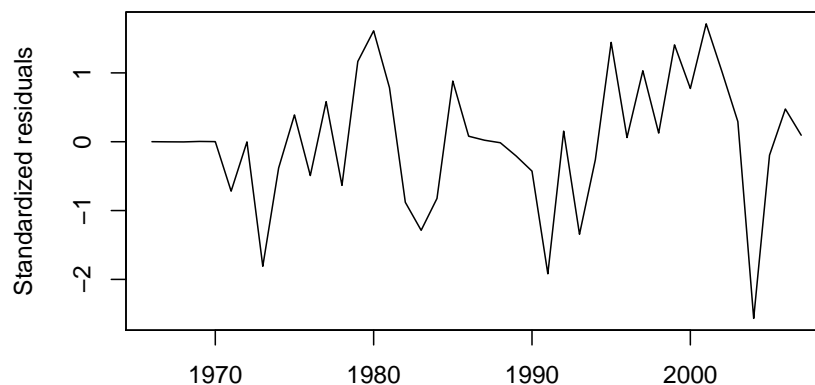
Figure A.3: Standardized residuals from model 2

Following Commandeur & Koopman, we present results in order of importance of assumptions: first independence (2007, pp. 90–96). Table A.4 shows results from the Box-Ljung test (see also, Petris, Petrone & Campagnoli 2009, pp. 93–95). The p-values reported indicate the significance of tests for autocorrelation of various lags of the standardized model residuals.

Statistic	p-value
$Q(1)$	0.1382
$Q(2)$	0.2907
$Q(3)$	0.3387
$Q(4)$	0.4055
$Q(5)$	0.5487
$Q(6)$	0.585
$Q(7)$	0.6378
$Q(8)$	0.6051
$Q(9)$	0.61
$Q(10)$	0.6994
$Q(11)$	0.7761
$Q(12)$	0.7141
$Q(13)$	0.7773
$Q(14)$	0.8284
$Q(15)$	0.8617

Table A.4: Box-Ljung tests for independence of standardized model residuals

The p-values are all greater than .05, which confirms with a high degree of confidence that the standardized residuals are uncorrelated.

Turning to the homoscedasticity assumption, we examine this following Commandeur & Koopman (2007, p. 92) by calculating the following statistic:

$$H(h) = \frac{\sum_{t=n-h+1}^{n} e_t^2}{\sum_{t=d+1}^{d+h} e_t^2}$$

Here, e_t indicates the standardized residual at time t, d is the number of diffuse initial elements (i.e. starting values of variances, of which there are two in our model), n is the length of the time series, and h is the integer nearest $(n-d)/3$. The test statistic of 1.736 is checked against the critical value of a two-tailed F-test with both degrees of freedom set to h. For test statistics greater than one (like ours), we compare it to the upper 0.025 critical level of the respective F distribution. Test statistics below one compare the reciprocal of the test statistic to the upper 0.025 critical level of the respective F distribution. In our case, the critical value of 3.474 is higher than our test statistic and therefore we do not reject the

null hypothesis of homoscedasticity.

Finally, we examine the normality of standardized residuals. The Shapiro-Wilk normality test returns a p-value of 0.1908, indicating that we cannot reject the null hypothesis of normally distributed residuals and supporting this assumption in our model.

E   Appendix: In-sample predictive accuracy

Given the results reported for the ability of the DLM to estimate dynamic or static coefficients accurately, it is comes as little surprise that the model's in-sample predictive accuracy is quite strong. Figures A.4 and A.5 plot actual standardized values of percentage change in U.S. defense spending against predictions from the two model specifications we present in the paper.
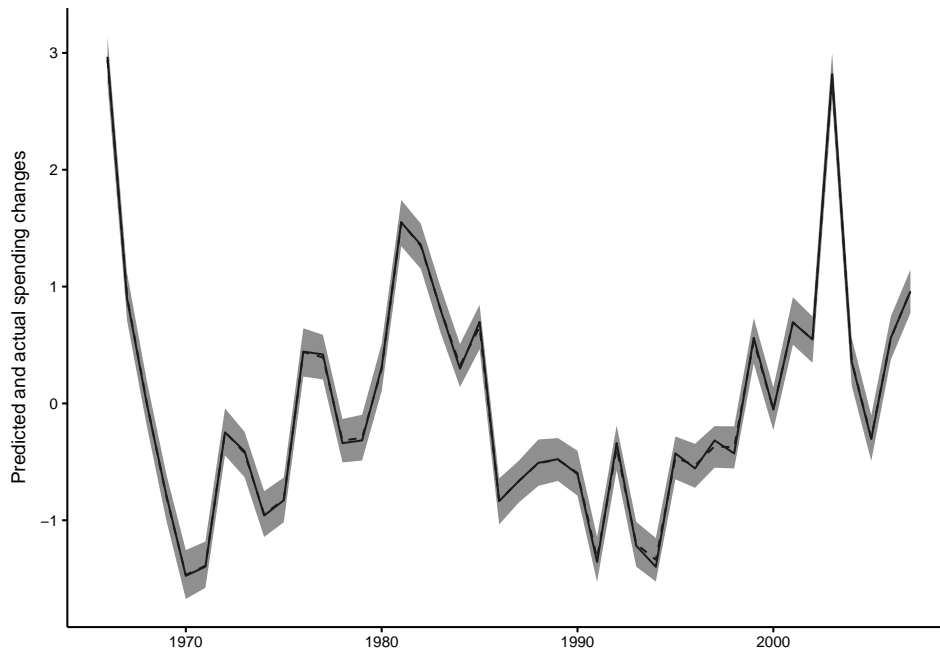
Figure A.4: In-sample predictive performance, model one

Note: The solid line is the actual (scaled) value of change in the defense budget. The dashed line plots central predictions from the model, and the gray region are simulated 95% confidence intervals.

In both figures, the actual values of the outcome are plotted with solid lines and the predictions from the model are plotted with dashed lines. Simulated confidence intervals are in gray. Note that the actual values of the outcome variable in figures A.4 and A.5 appear different because they are standardized by subtracting the mean of the time series and dividing by its standard deviation. The standardized values differ because the second model uses a slightly longer time series. Overall, we take this strong predictive accuracy as a sign that our model specification is appropriate to the data and that the models are

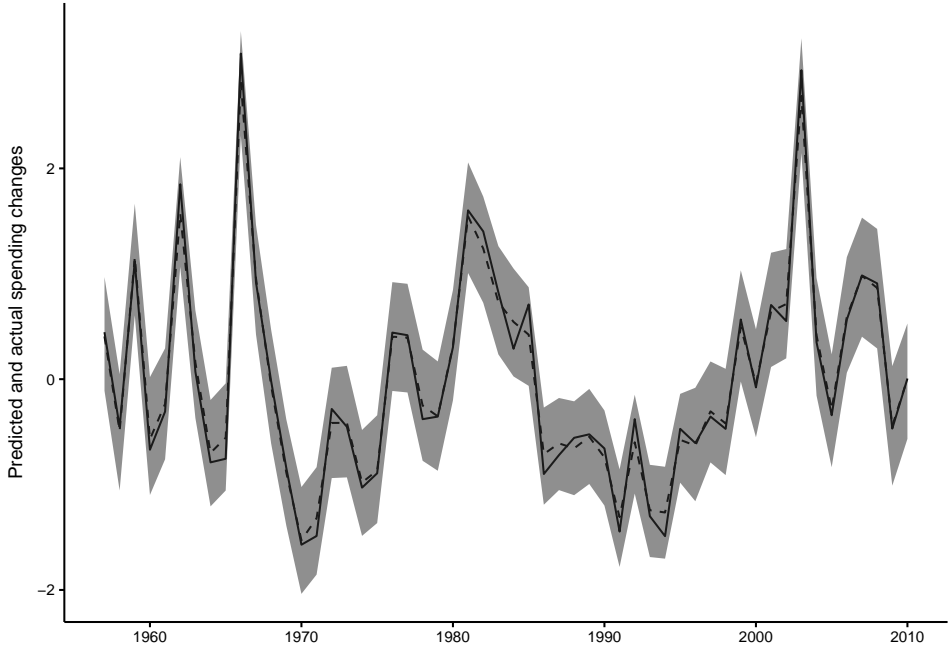capturing substantively informative changes in coefficient estimates over time.

Figure A.5: In-sample predictive performance, model two

Note: The solid line is the actual (scaled) value of change in the defense budget. The dashed line plots central predictions from the model, and the gray region are simulated 95% confidence intervals.

F Appendix: Replicating True's original time frame

This section presents DLM results for a regression using the exact time frame (1966–1992) and the same independent variables as in True's time-varying analysis (see True 2002). We have repeated the analysis twice. First, we use both True's exact measure of the dependent variable — dollar amount annual changes in U.S. defense spending — and his measure of lagged spending, which is a simple lag of the dollar amount annual changes in U.S. defense spending. Secondly, we run the analysis again with our preferred measures: percentage annual changes in U.S. defense spending as dependent variable and the lagged level of real defense spending. As in the main text, our findings differ somewhat from those reported by True in both cases.

Time-varying coefficient estimates with confidence intervals for the dollar amount changes reported by True are plotted in figure A.6. For one, we find a systematic positive relationship between periods of war or international tension and changes in U.S. defense spending only during the Reagan buildup and the wind-down in defense spending following the Vietnam war. The relatively small drop-offs in defense spending following the end of the Gulf War flip this association to an apparently negative one. This is a result of considering the time series as a whole, rather than artificially subsetting it into epochs. As noted in the main text, the DLM allows us to detect changes in the effects of war or tension over time, therefore the most appropriate way for us to examine True's ideas is to combine the three dummy variables for Vietnam, the Gulf War, and the Reagan buildup into a single dummy variable.

International support (aid), on the other hand, exhibits a pattern that resembles True's findings rather well. The confidence intervals in figure A.6 are difficult to discern, but the association between international aid and actual changes in U.S. defense spending is positive and significant in the mid-1980s, as True also finds. Soviet spending is also estimated to have an effect in line with True's findings, though we identify a positive association with actual changes in U.S. defense spending and total Soviet defense spending to exist throughout the period, once we account for the entire time series.
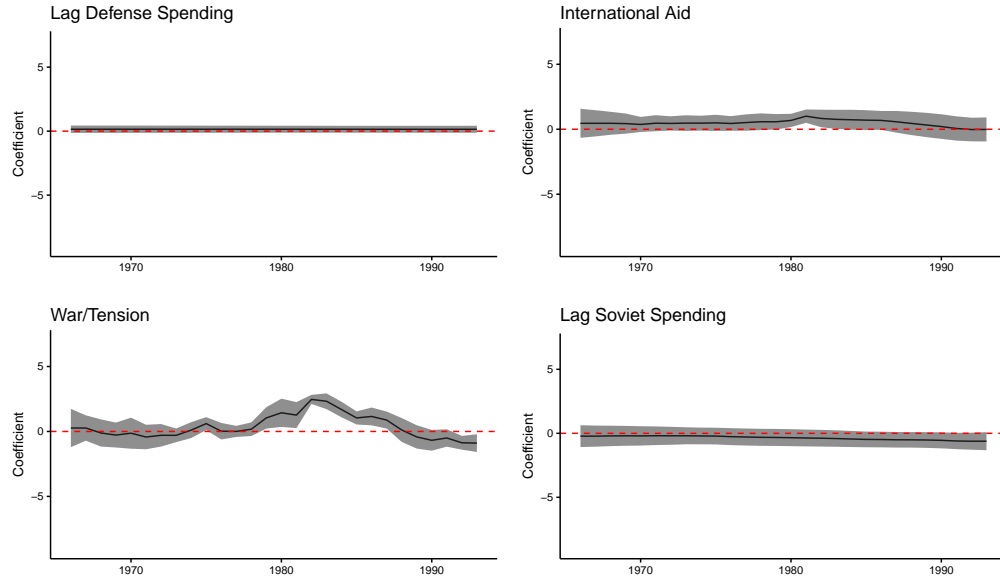
Figure A.6: Estimated time-varying coefficients as in True's table 7.3
(DV: actual dollar changes in defense spending)

Note: Effect coefficients are plotted over time. The solid line indicates the point estimate for the coefficient each period, while gray areas are 95% confidence intervals. The dashed horizontal line indicates zero. Where the gray area overlaps the dashed line, estimates are not statistically significant by conventional standards.

Contrary to True, however, our model using his exact time period and measurement choices indicates no systematic relationship between either the lagged dollar amount change in U.S. defense spending or the lagged level of Soviet defense spending at any point in the time series.

Turning to our preferred way of measuring the dependent variable and lagged spending, in figure A.7 we find a different story entirely. First of all, once we measure changes in U.S. defense spending as the percentage change from the previous year and treat lagged spending as the actual previous level of spending, we find no statistically discernible relationship between periods of war or tension and the brief significant relationship between international aid and changes in defense spending is now reversed. The first finding we attribute to the benefit of considering the entire time series: the relationships identified by True are too weak in terms of the percentage of the budget and too inconsistently related to periods of war or tension to be attributed to the events themselves. Likewise, the finding regarding

A15

international aid looks quite different when asking about aid's relationship to the percentage change in the defense budget instead of the actual dollar amount changes in the defense budget.
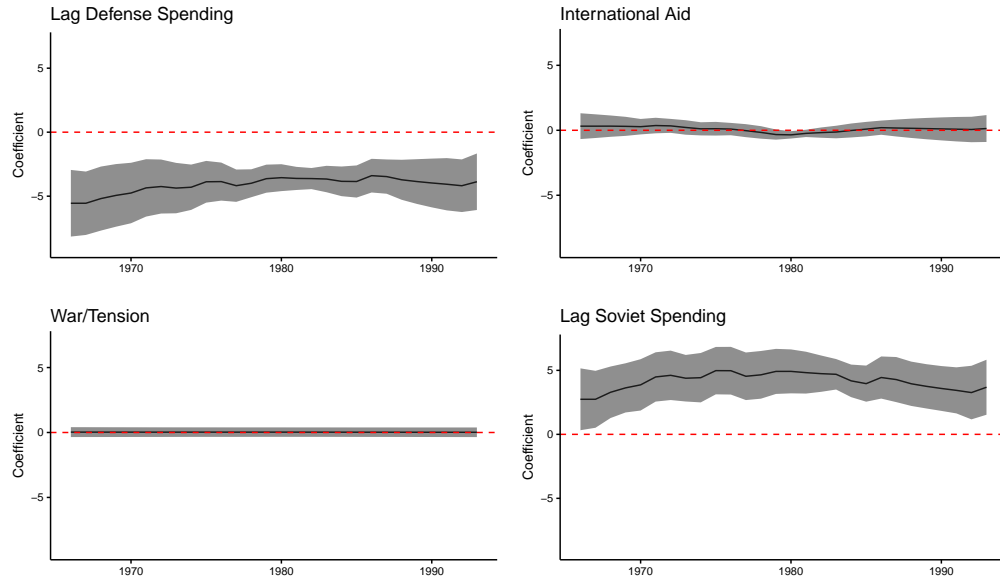

Figure A.7: Estimated time-varying coefficients as in True's table 7.3
(DV: percentage changes in defense spending)

Note: Effect coefficients are plotted over time. The solid line indicates the point estimate for the coefficient each period, while gray areas are 95% confidence intervals. The dashed horizontal line indicates zero. Where the gray area overlaps the dashed line, estimates are not statistically significant by conventional standards.

Finally, we find that the lagged level of U.S. defense spending is consistently negatively related to percentage changes in the defense budget, and the lagged level of Soviet defense spending is consistently positively related to percentage changes in the U.S. defense budget. The former finding is perhaps surprising, but, as in the main text, it is due to the fact that the data are mean-centered and divided by their standard deviation before analysis. This means that values above the mean are positive, and values below the mean are negative. Substantively, this finding means that larger than average U.S. defense budgets are likelier to be cut and smaller than average defense budgets are likelier to increase — and the greater in magnitude these levels are above or below the mean, the larger the resulting cut or increase is expected to be. The same story applies in regard to the level of Soviet military spending:

when Soviet spending was above average, U.S. defense spending was likely to experience an increase and when Soviet spending was below its average U.S. spending was more likely to be cut.

G Appendix: Simulating data of various lengths

This section contains the results of simulations similar to those presented in the main text, but using time periods of varying lengths. The tests presented here vary from 30 to 100 time periods — i.e. these are tests conducted on randomly generated data to demonstrate that the DLM can recover estimates from data sets of the respective length of time and do not represent real empirical data.
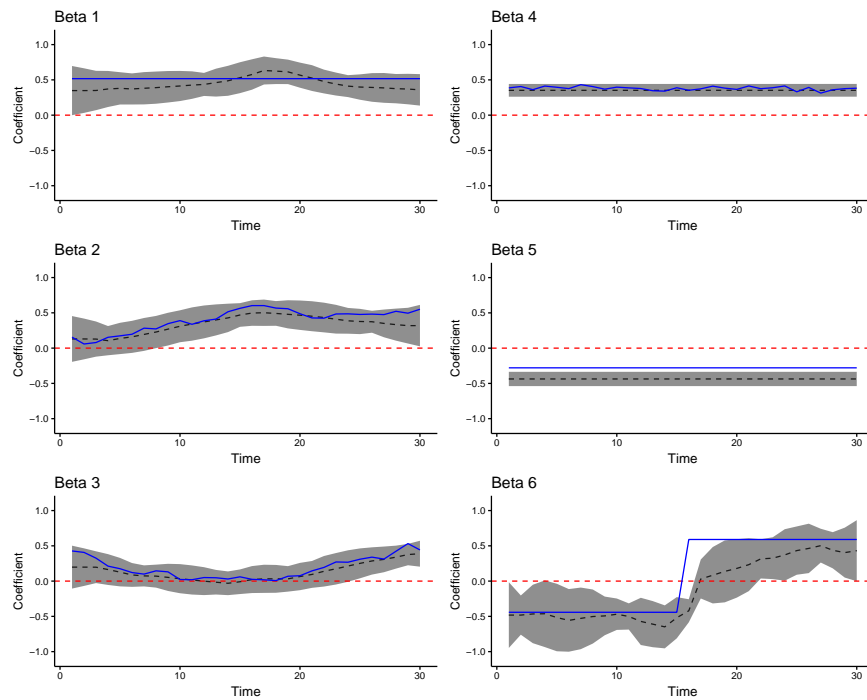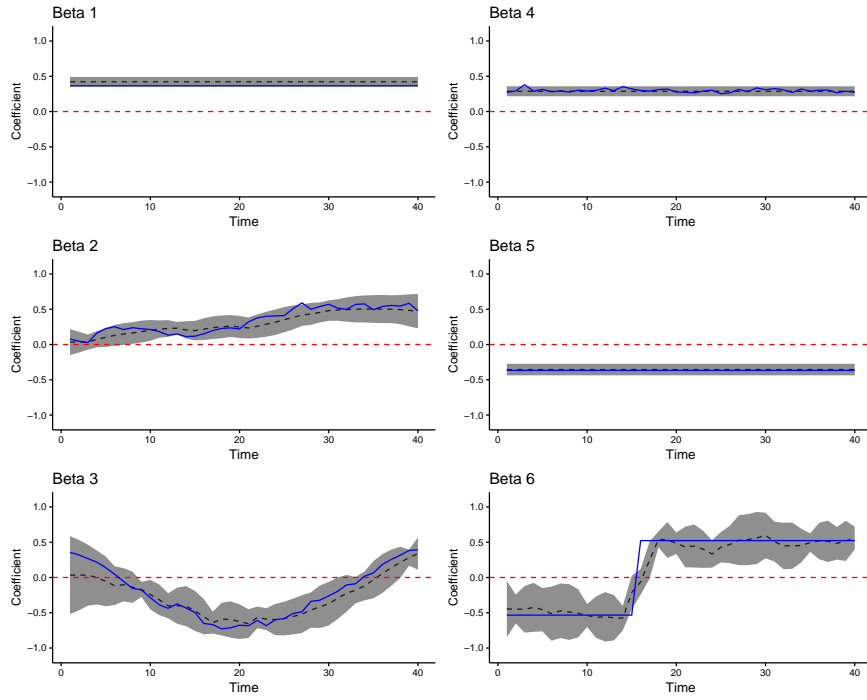

Figure A.8: Simulated results with 30 time periods

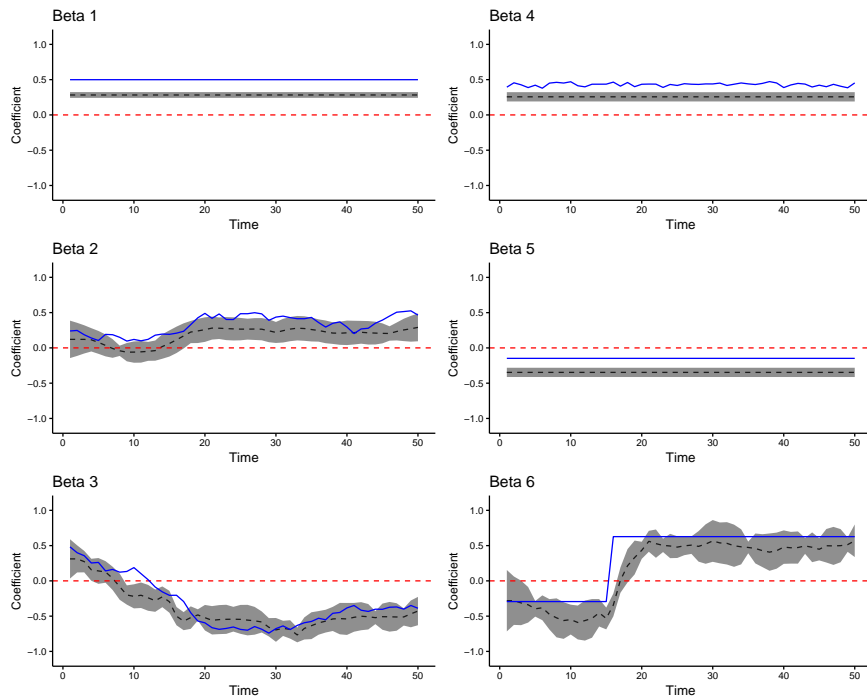Figure A.9: Simulated results with 40 time periods

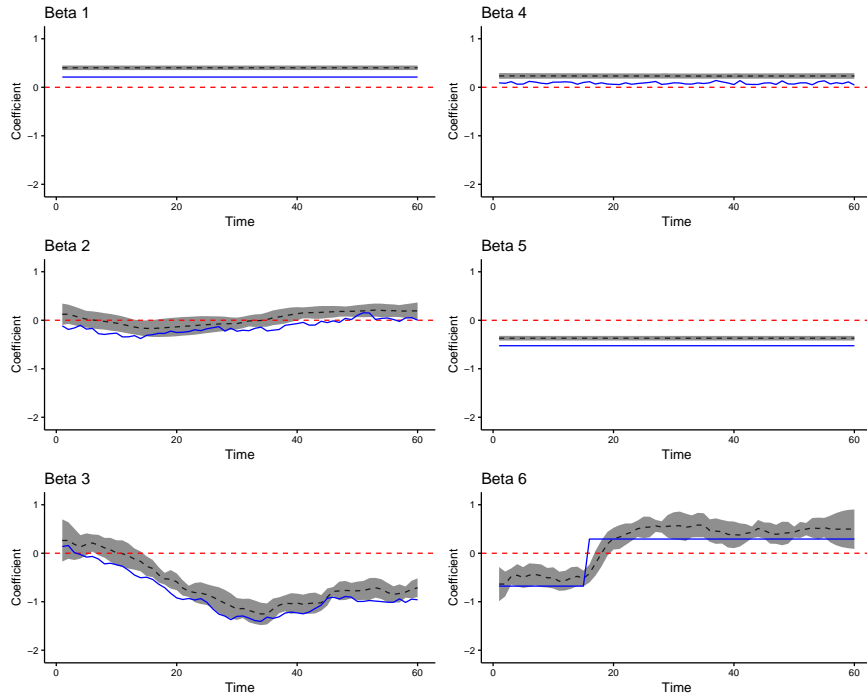Figure A.10: Simulated results with 50 time periods

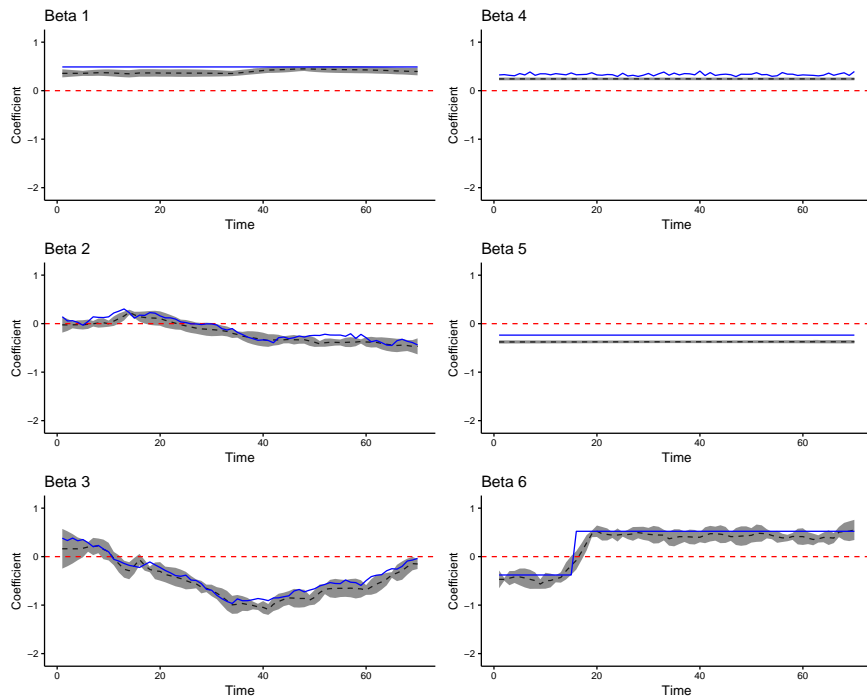Figure A.11: Simulated results with 60 time periods

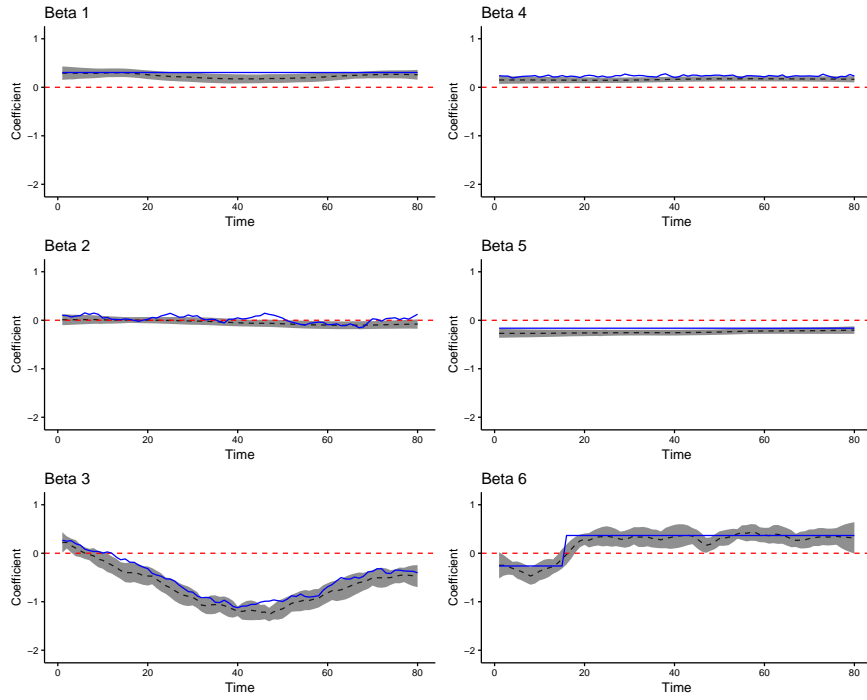Figure A.12: Simulated results with 70 time periods

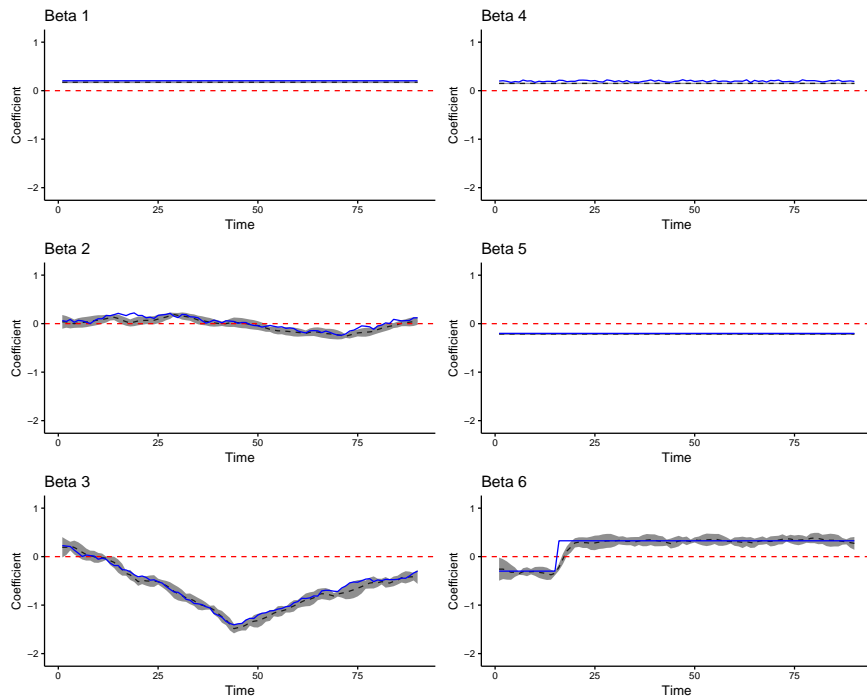Figure A.13: Simulated results with 80 time periods

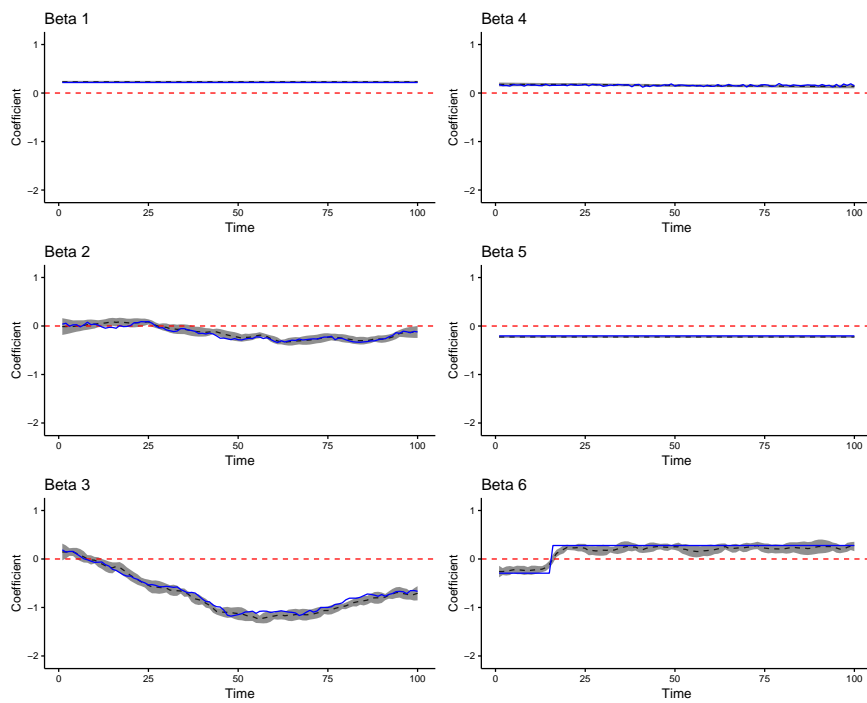Figure A.14: Simulated results with 90 time periods

Figure A.15: Simulated results with 100 time periods

H Appendix: Sample simulation code

Below is sample R code for running a simulation of the analyst's own creation. We recommend altering the parameters and model structure in this code as needed to build simulations that demonstrate the feasibility of a DLM recovering the "true" parameters in real data of the same length and with the same type of (expected) dynamics in the analyst's own data.

```r
library(ggplot2) #(version 2.2.1)
library(dlm) #(version 1.1-4)
library(MASS) #(version 7.3-47)
library(zoo) #(version 1.8-0)
library(data.table) #(version 1.10.4)

# Set random seed to ensure replicability
set.seed(9715)

# Set length of simulated time series
length.of.series <- 60

# Simulate true time-varying effects
   ## Beta 1 = randomly determined constant (near 7)
   b1 <- rep(x = rnorm(n = 1, mean = 7, sd = 1),
                  times = length.of.series)

   ## Beta 2 = randomly evolving (period effect modified by standard normal draw)
   b2 <- rep(x = NA, times = length.of.series)
   b2[1] <- 1
   for (i in 2:length.of.series) {
      b2[i] <- b2[i-1] + rnorm(n = 1, mean = 0, sd = 1)
   }

   ## Beta 3 = Trend downward from 6 (on average) for the first half of the
   ## time series, then trend upward (on average) for the second half
   b3 <- rep(NA, length.of.series)
   b3[1] <- 6
   for (i in 2:length.of.series){
      if (i < length.of.series * .5) {
         b3[i] <- b3[i-1] + rnorm(n = 1, mean = -.9, sd = 1)
      } else {
         b3[i] <- b3[i-1] + rnorm(n = 1, mean = .5, sd = 1)
      }
```

```
    }

    ## Beta 4 = Gaussian noise variation over time around 5
    b4 <- rnorm(n = length.of.series, mean = 5, sd = .5)

    ## Beta 5 = Constant effect (equal to -7)
    b5 <- rep(-7, length.of.series)

    ## Beta 6 = Equal to -10 for 15 periods, then jumps to 9
    ## and stays constant
    b6 <- rep(-10, 15)
    b6 <- c(b6, rep(x = 9, times = (length.of.series - 15)))


# Group simulated betas into a matrix
B <- as.matrix(cbind(b1, b2, b3, b4, b5, b6))

# Generate X as a series of standard normally distributed random draws
x1 <- rnorm(length.of.series)
x2 <- rnorm(length.of.series)
x3 <- rnorm(length.of.series)
x4 <- rnorm(length.of.series)
x5 <- rnorm(length.of.series)
x6 <- rnorm(length.of.series)

# Group x's into a matrix
X <- as.matrix(cbind(x1, x2, x3, x4, x5, x6))

# Set simulated variance of y
sig2 <- 1

###################################################
# Calculate y as a linear function of x * beta
# and a random error term
    ## Calculate mean value of Y
    mu <- rowSums(B * X)

    ## Generate simulated error terms
    errors <- rnorm(n = length.of.series, mean = 0, sd = sqrt(sig2))

    ## Calculate y
    y <- mu + errors

# Generate time period (vector of times)
time.period <- 1:length.of.series

# Generate time series objects for analysis
data <- ts(X, start = min(time.period))
```

```
y <- ts(y, start = min(time.period))

###################################################
# Run the model
   ## Function to build the DLM
   buildTVP <- function(u) {
      dlmModReg(X, addInt = F, dV = exp(u[1]),
                    dW = exp(u[2:num]),
                    m0 = rep(0, (num - 1)),
                    C0 = 1e+07 * diag((num - 1)))
   }

   ## Pick right dimension of priors for size of the X matrix
   num <- NCOL(X)+1

   ## Estimate DLM starting values
   outMLE <- dlmMLE(scale(y), parm = rep(0, num), buildTVP, hessian = T,
                    method = "BFGS")

   ## Build smoothing model
   mod <- buildTVP(outMLE$par)

   ## Run filter
   filtered <- dlmFilter(scale(y), mod)

   ## Run smoother on filtered results
   results <- dlmSmooth(filtered)

###################################################
# Set up labels and data matrix of results
   ## Add one to time period to accomodate a single coefficient forecast
   ## from filtering stage (this gets dropped when plotting)
   time.period.est <- c(time.period, (max(time.period) + 1))

   ## Extract smoothed parameter estimates
   param <- results$s
   colnames(param) <- c(paste0("beta", 1:ncol(param)))

   ##Store parameter labels for matrix to send to plot
   labels <- rep(colnames(param), nrow(param))

   ##Generate data matrix to send to graph
   to.graf <- as.data.frame(cbind(labels[order(labels)],
                                  as.vector(param),
                                  rep(x = time.period.est, times = ncol(param))),
                     stringsAsFactors= F)

   ## Column names for data matrix to plot
```

```
    names(to.graf) <- c("param", "value", "year")


# Calculate and format smoothed confidence intervals
   ## Calculate variance-covariance matrix of parameters
   ## at each discrete time point
   vc.mats <- dlmSvd2var(results$U.S, results$D.S)

   ## Initialize empty list for confidence interval data matrices
   intervals <- list()

   ## Extract confidence 95% intervals for coefficient
   ## estimates at each discrete time point
   for (i in 1:nrow(param)) {
      intervals[[i]] <- as.data.frame(cbind(colnames(param),
                              param[i,] - 1.96 * sqrt(diag(vc.mats[[i]])),
                              param[i,] + 1.96 * sqrt(diag(vc.mats[[i]])),
                              rep(x = time.period.est[i], times = ncol(param))),
                        stringsAsFactors = F)
   }

   ## Munge the confidence intervals into a plot-ready format
   cis <- as.data.frame(rbindlist(intervals))
   names(cis) <- c("param", "ci.lo", "ci.hi", "year")

   ## Merge together confidence intervals and point estimates
   final.to.graf <- merge(to.graf, cis)
   final.to.graf[, 2:5] <- apply(final.to.graf[, 2:5], 2, as.numeric)

####################################################
# Generate separate coefficient plots
   ## Generate plot labels for coefficient graphs
   labs <- c("Beta 1","Beta 2","Beta 3","Beta 4","Beta 5","Beta 6")

   ## Remove final forecast of coefficients
   params <- final.to.graf[final.to.graf$year != max(time.period.est), ]

   ## Split parameter estimates plus confidence intervals
   ## by explanatory variable
   sep <- split(params, params$param)

   ## Extract column titles for later
   column.titles <- colnames(sep[[1]])

   ## Place 'true' simulated betas on same scale as coefficient
   ## estimates from the model estimated on standardized y
   B.for.plot <- (B - mean(y)) / sd(y)
```

A26

```
## Produce graphs
## Loop over explanatory variables, plotting each one
for (i in 1:length(sep)) {
   #Reorder data for printing by time period variable:  "year"
   sep[[i]] <- sep[[i]][order(sep[[i]]$year), ]

   # Produce matrix of estimates and true beta
   out <- cbind(sep[[i]], B.for.plot[,i])

   # Name columns in that matrix
   names(out) <- c(column.titles, "actual")

   # Set plot parameters
   y.hi <- 1.2 #(upper limit of plot range)
   y.lo <- -1 #(lower limit of plot range)

   # Print plot to screen
   ggplot(out, aes(x = year)) +
      geom_line(aes(y = value), linetype = "dashed") +
      geom_ribbon(aes(ymin = ci.lo, ymax = ci.hi, alpha = .001)) +
      scale_y_continuous("Coefficient", limits = c(y.lo, y.hi)) +
      scale_x_continuous("Time") +
      theme(panel.background = element_blank(), legend.position = "none",
               axis.line = element_line(colour = "black")) +
      ggtitle(labs[i]) +
      geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
      geom_line(aes(y = actual), col = "blue")

   # Save graph for later
   # (Uncomment line to run)
   # ggsave(file = paste0("sim_beta_", i, ".pdf"), width = 5, height = 3)
}
```

I Appendix: Main results without smoothing

This section presents the results of our main analysis excluding the Kalman smoothing step. As we have mentioned, the Kalman smoothing step follows filtering and uses a backward-recursion to condition each period's parameter estimates on the conditional estimates of the parameters at time $t + 1$ (Petris, Petrone & Campagnoli 2009, Shumway & Stoffer 2010). The variances of each time period's coefficient estimates tend to be very large earlier in the time series for filtered models, because filtered estimates are conditioned only on the first t observations. If t is small, i.e. earlier in the time series, then uncertainty is relatively large. The smoothed estimates have less uncertainty because they condition each period's coefficient estimates on the full time series. Conditioning on the full time series can also have the effect of "smoothing" out coefficient estimates that are overly large, given all of the information in the data, pushing dynamic coefficient estimates toward displaying less change over time.

Choose to apply smoothing, versus simply filtering a DLM, is a choice that many recommend be made depending on the purpose of the model (See, for example Petris, Petrone & Campagnoli 2009). Filtering only is the most effective approach for forecasting future outcomes from time series. However, to examine dynamic evolution of the underlying states - i.e. our coefficient estimates - in a context in which we are interested in the political process underlying the data, smoothing is generally best because it reexamines each filtered estimate using all of the information in the time series.

We argue, based on this reasoning, that applying a smoother to our models is the right decision. However, we present the filtering-only results in this section to address the concern that the filtered and smoothed results look very different from one another. If the smoothed results were dramatically dampening the shape of the dynamic estimates from the filter, then we would be concerned that our final model understates the evidence for dynamic relationships in our data.

Figure A.16 shows the results from our main model (model 2 in the main text) using only

the filtered model, before applying a smoother. Note that the time series over which we have plotted the results is shorter by around ten years than that plotted in figure 2. The model still includes those observations, but we have not plotted them here because the confidence intervals are massive - extending above 5,000 and below -5000. As we would expect, these confidence intervals shrink rapidly as the filtering recursion proceeds through the data, so that we can plot results from around the mid-1960s and the graphs remain readable.
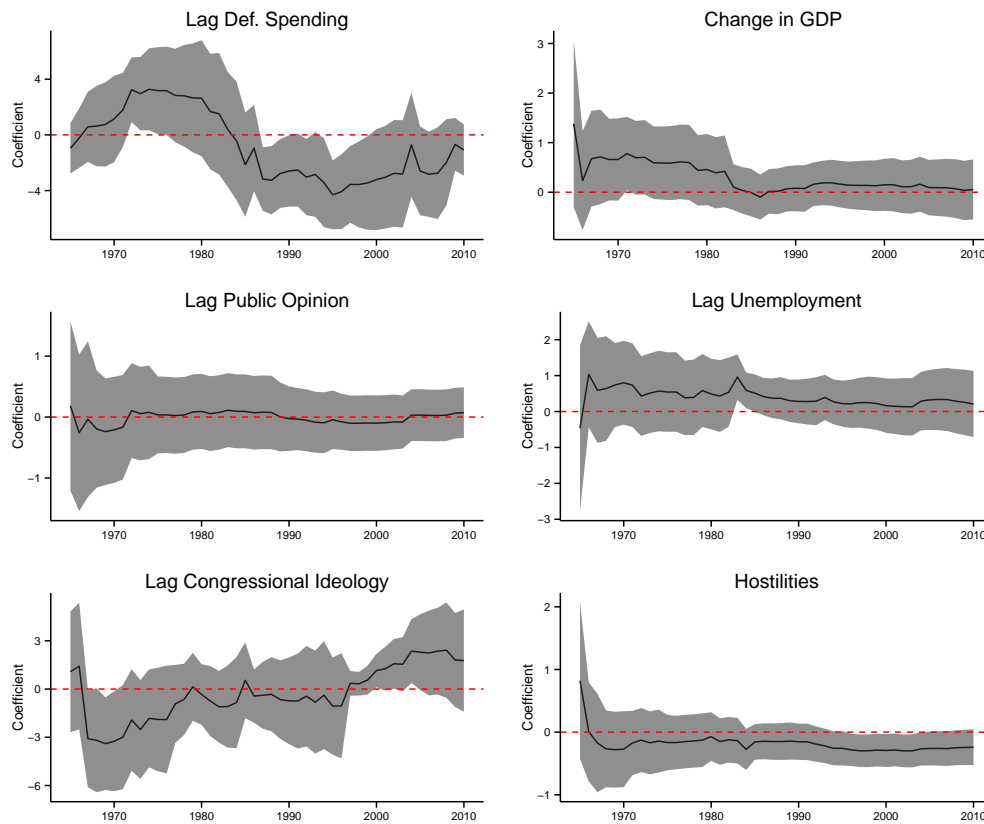

Figure A.16: Estimated time-varying coefficients - filtered only

Note: Effect coefficients are plotted over time. The solid line indicates the point estimate for the coefficient each period, while gray areas are 95% confidence intervals. The dashed horizontal line indicates zero. Where the gray area overlaps the dashed line, estimates are not statistically significant by conventional standards.