A Dynamic Linear Modeling Approach to Public Policy Change

Matt W. Loftis*
Assistant Professor
Dept. of Political Science
Aarhus University
Telephone: +45 8716 5635
Fax: +45 8613 9839
mattwloftis@ps.au.dk

Peter B. Mortensen
Professor
Dept. of Political Science
Aarhus University
Telephone: +45 8716 5707
Fax: +45 8613 9839
peter@ps.au.dk

Keywords: punctuated equilibrium theory, public policy, policy dynamics, public spending

Wordcount: $\approx$ 7,500

* Working paper, please do not cite without permission. The authors wish to thank Gregory McAvoy, participants at the Aarhus University Department of Political Science Public Administration workshop, and panel participants at the Comparative Agendas Conference 2015 and Midwest Political Science Association general meeting 2015. Replication materials are available at:
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/H95OL1 &version=DRAFT

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:

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

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

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 $t-1$: $y_{t|t-1} = \boldsymbol{X_t \beta_{t|t-1}}$

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

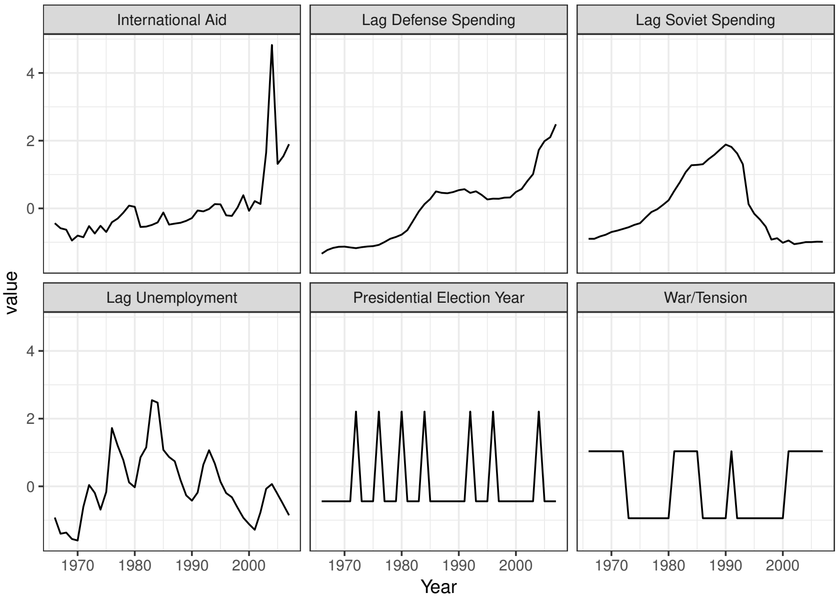

Figure A.1: Standardized variables from model 1

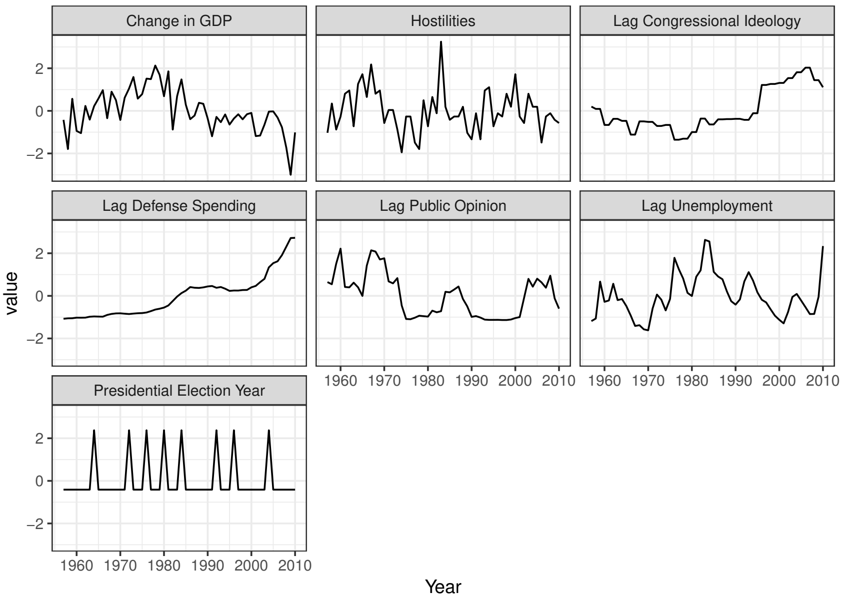

Figure A.2: Standardized variables from model 2

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.

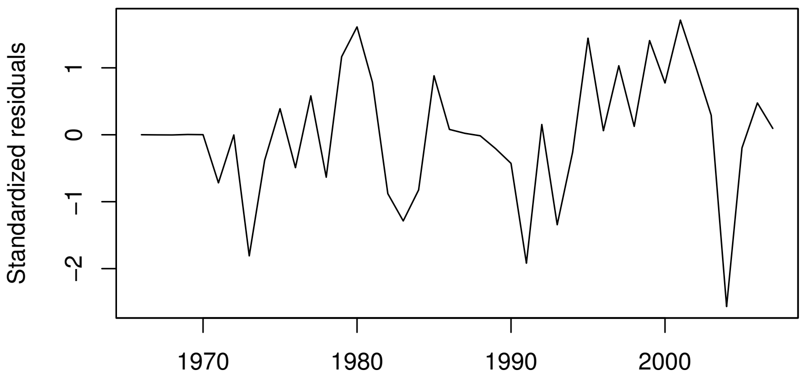

Figure A.3: Standardized residuals from model 2

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.

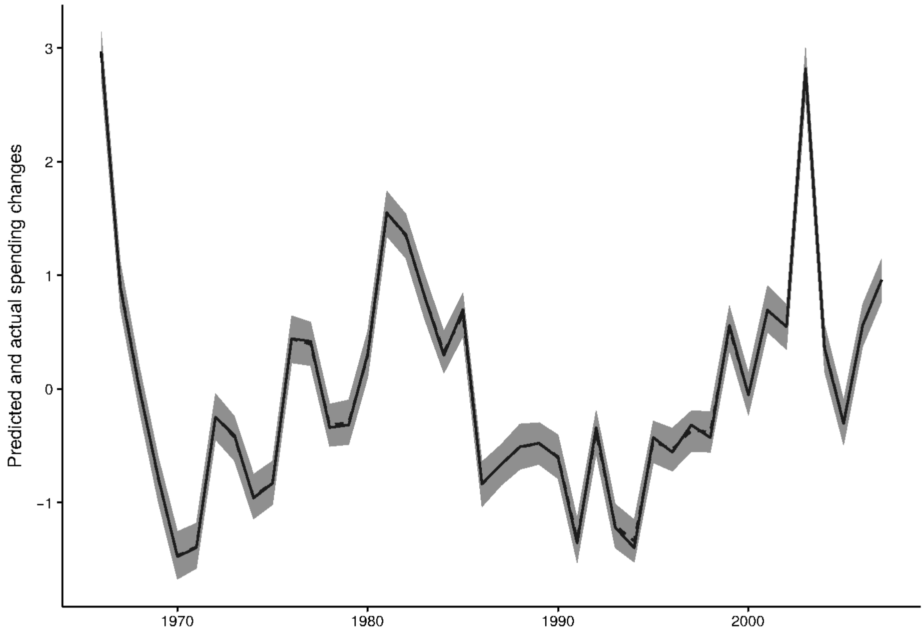

Figure A.4: In-sample predictive performance, model one

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.

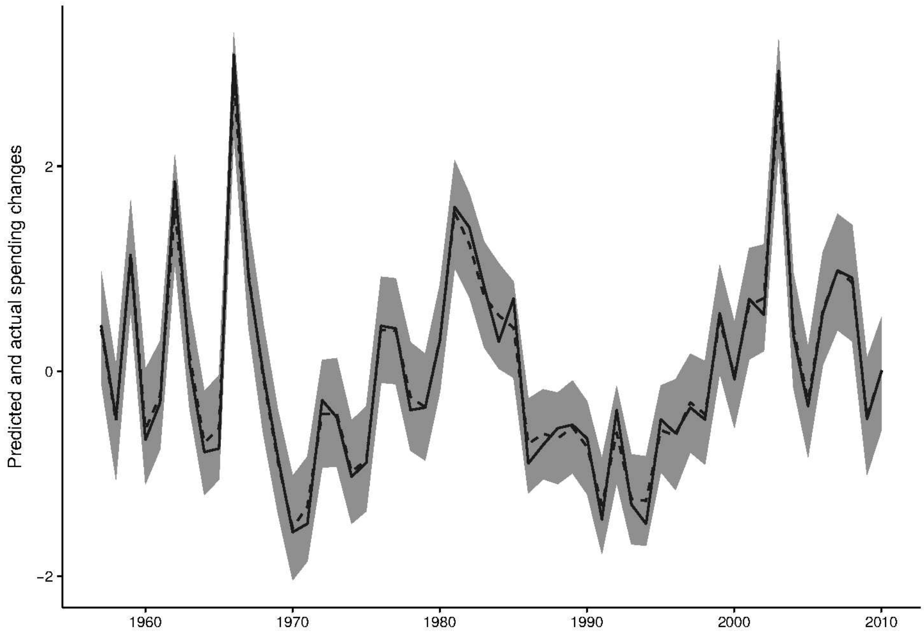

Figure A.5: In-sample predictive performance, model two

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.

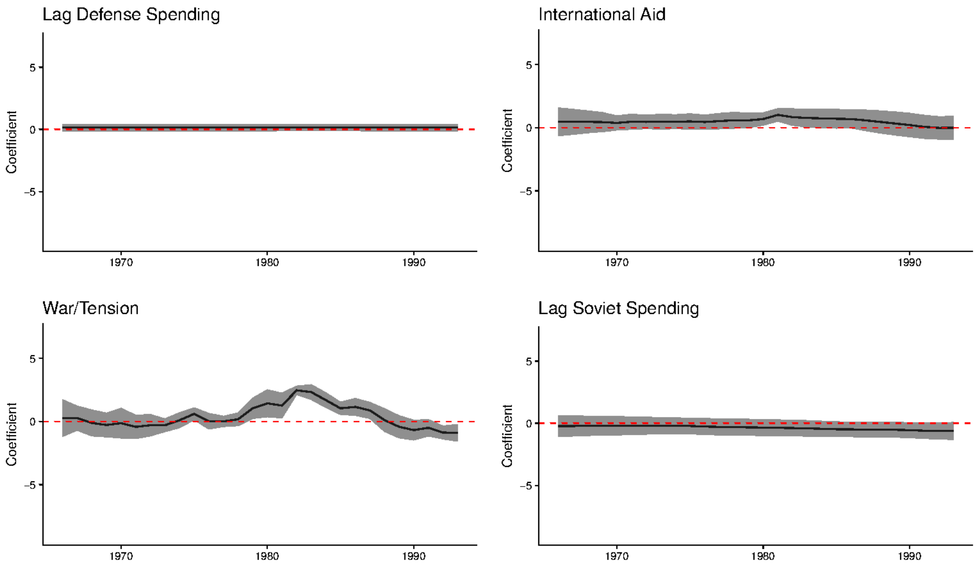

Figure A.6: Estimated time-varying coefficients as in True's table 7.3
(DV: actual dollar changes in defense spending)

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 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.

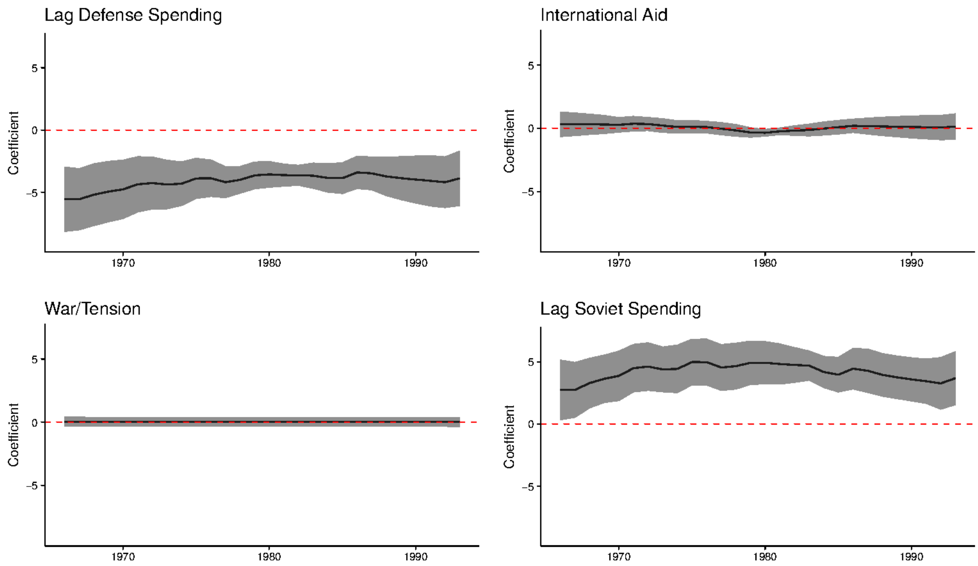

Figure A.7: Estimated time-varying coefficients as in True's table 7.3
(DV: percentage changes in defense spending)

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.

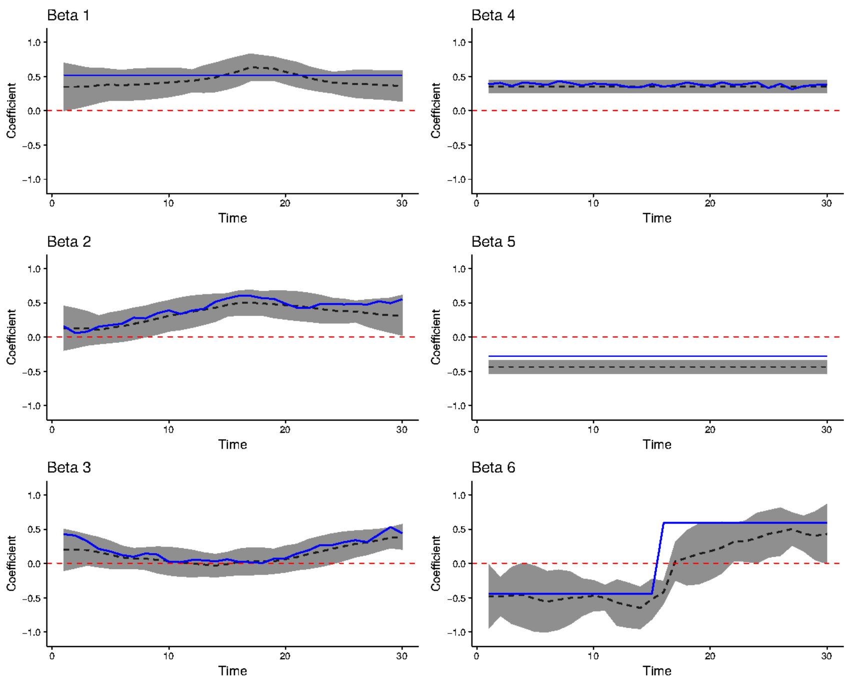

Figure A.8: Simulated results with 30 time periods

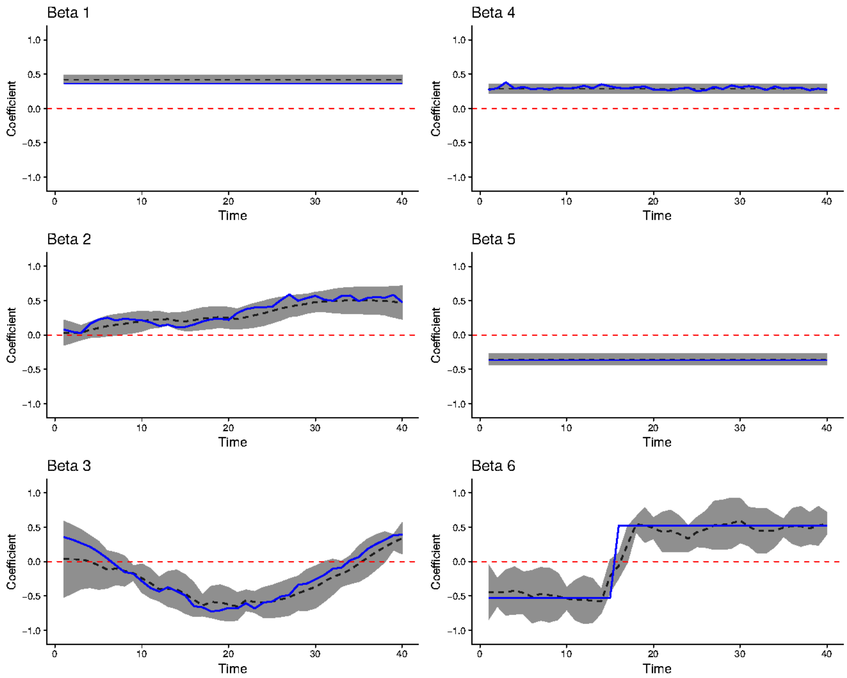

Figure A.9: Simulated results with 40 time periods

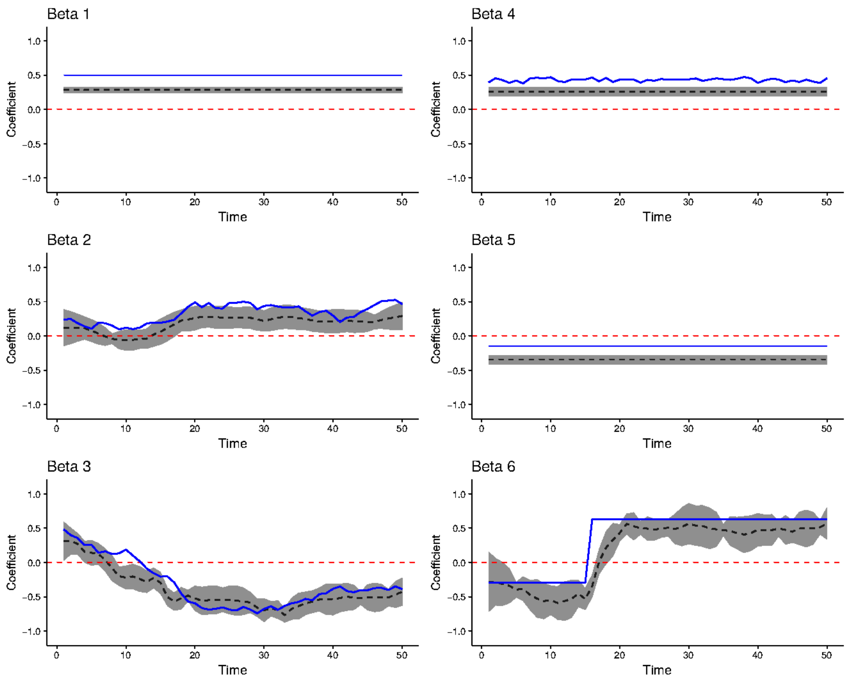

Figure A.10: Simulated results with 50 time periods

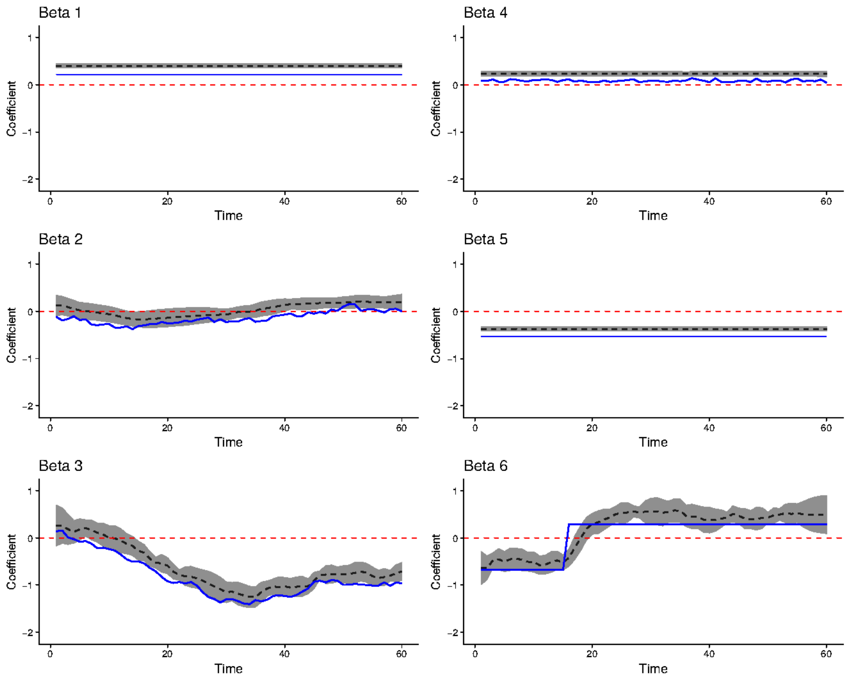

Figure A.11: Simulated results with 60 time periods

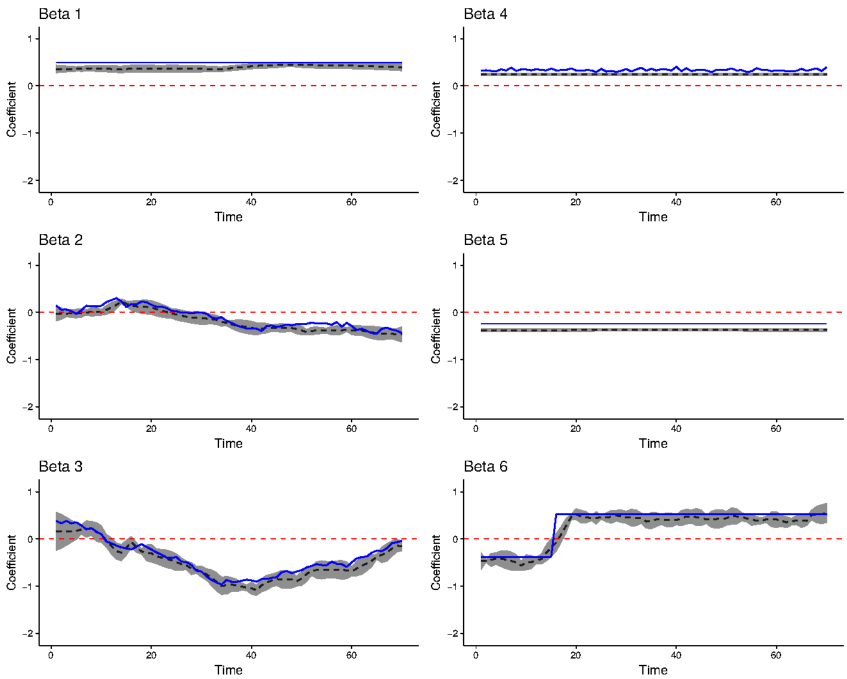

Figure A.12: Simulated results with 70 time periods

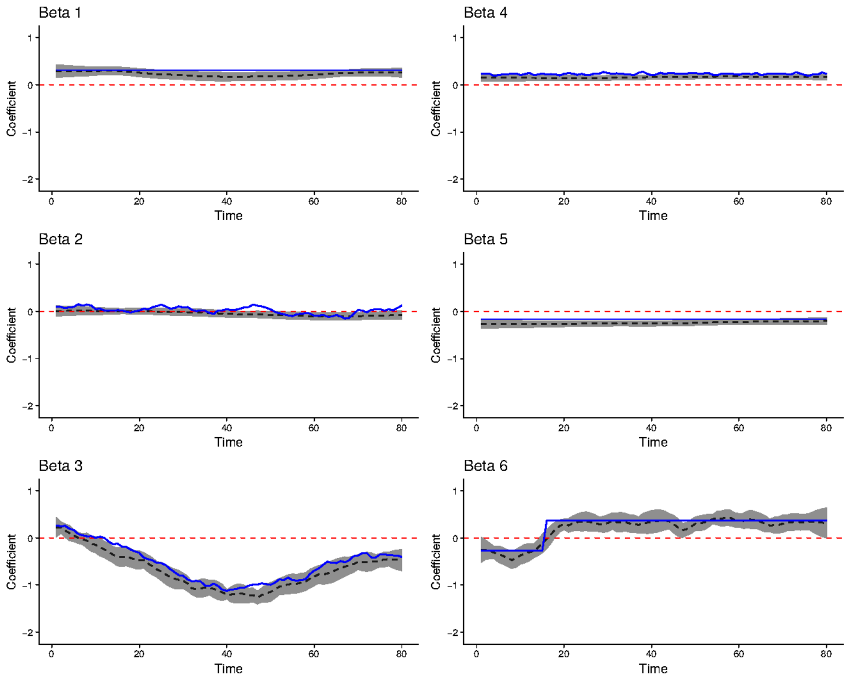

Figure A.13: Simulated results with 80 time periods

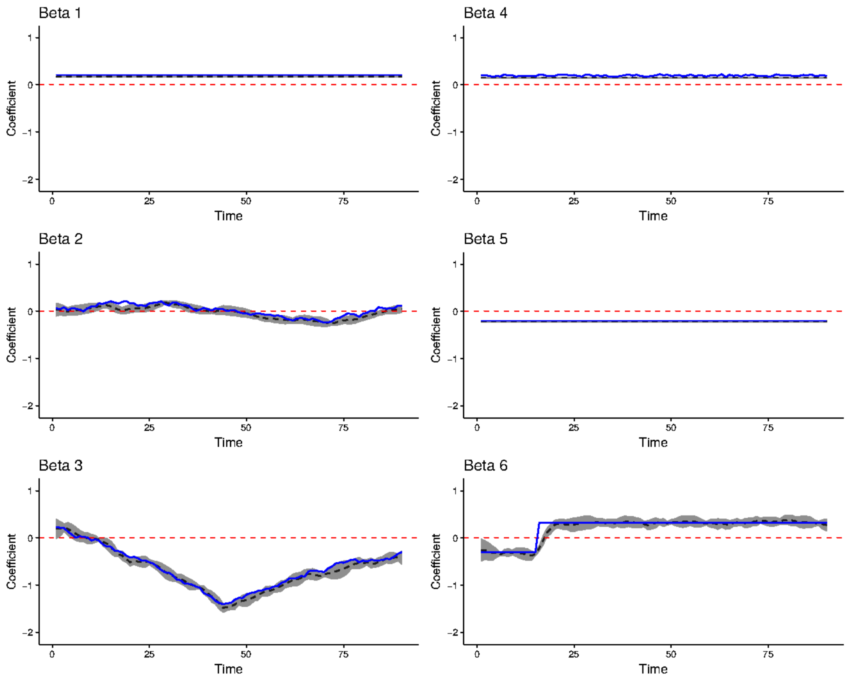

Figure A.14: Simulated results with 90 time periods

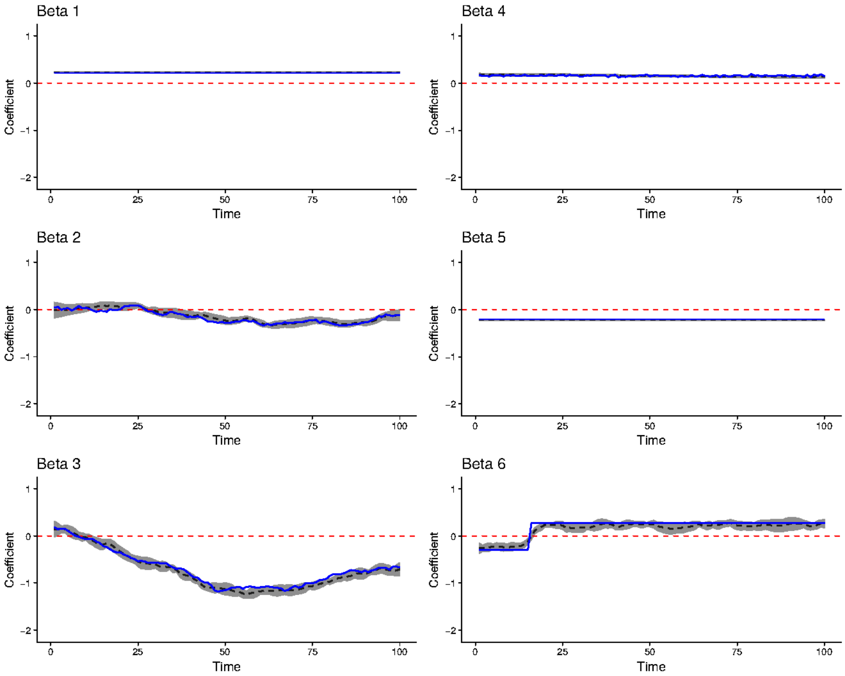

Figure A.15: Simulated results with 100 time periods

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.

```
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)
```

```
## 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.

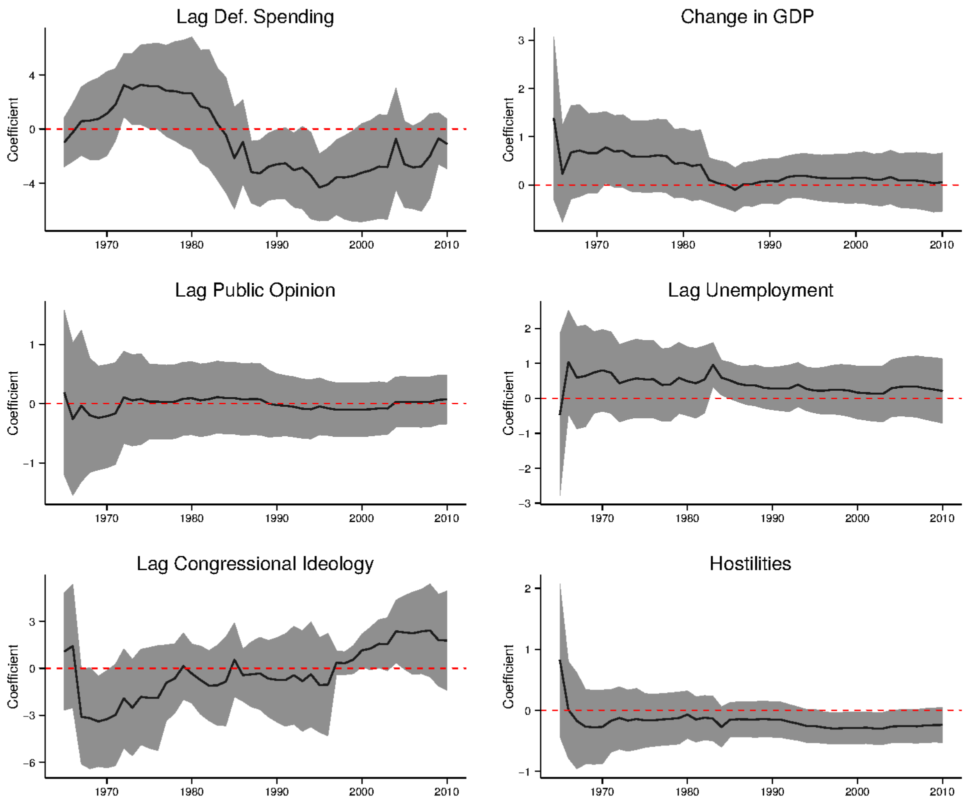

Figure A.16: Estimated time-varying coefficients - filtered only

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.