Finding Jumps in Otherwise Smooth Curves: Identifying Critical Events in Political Processes

Marc T. Ratkovic

*Departments of Political Science and Statistics, University of Wisconsin, Madison,
110 North Hall, 1050 Bascom Mall, Madison, WI 53706*

Kevin H. Eng

*Department of Statistics, University of Wisconsin, Madison, 1300 University Ave.,
Madison, WI 53706*
e-mail: ratkovic@wisc.edu (corresponding author)

Many social processes are stable and smooth in general, with discrete jumps. We develop a sequential segmentation spline method that can identify both the location and the number of discontinuities in a series of observations with a time component, while fitting a smooth spline between jumps, using a modified Bayesian Information Criterion statistic as a stopping rule. We explore the method in a large-*n*, unbalanced panel setting with George W. Bush's approval data, a small-*n* time series with median DW-NOMINATE scores for each Congress over time, and a series of simulations. We compare the method to several extant smoothers, and the method performs favorably in terms of visual inspection, residual properties, and event detection. Finally, we discuss extensions of the method.

1 Introduction

Political processes are generally stable and smooth but occasionally explosive around critical events (Pierson 2004). Simultaneously identifying the jumps and estimating a smooth curve pose a particular statistical challenge. In this article, we adapt a smoothing spline in a manner that allows identification of both the location and the number of breaks in a time series, producing a fitted function that represents both critical and secular change.

Our interest in this problem began when we noticed the ''9/11 problem'' while trying to fit smoothers to George W. Bush's approval ratings, as illustrated in Fig. 1. Smoothers, to some extent or another, show an uptick in Bush's approval well before 9/11. The spline in Fig. 1 would lead to the conclusion that the increase in approval began in mid August, which is clearly incorrect. Bush's approval was slowly trending downward until 9/10, and then by 9/12 it was up around 80% or even higher. Before 9/11, the spline estimate systematically overestimates Bush's approval, while systematically underestimating Bush's approval post-9/11.
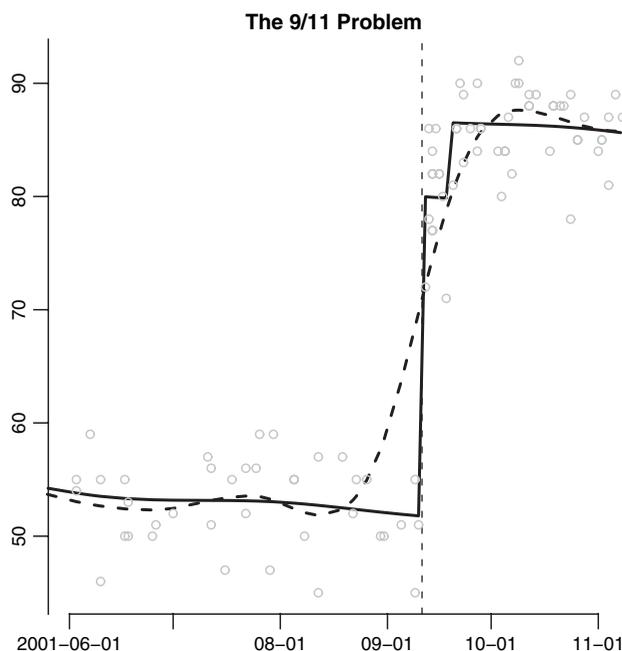
Fig. 1 Smoothing splines are dashed and our method is solid. The smoothing splines show an uptick in Bush's popularity in mid-July, well before 9/11.

Similarly, when using a loess smoother, optimal choice of the tuning parameter resulted in a curve nearly identical to the generalized smoothing spline, also missing 9/11. Decreasing the span until it picks up the 9/11 jump creates too much variance elsewhere. From a time series standpoint, although presidential approval may be described well *on average* as an autoregressive process (Erikson, MacKuen, and Stimson 2002), it is not *everywhere* the same process. The "memory" in the process immediately post-9/11 was rather short, whereas during more mundane times, the memory may be longer. Nathaniel Beck and colleagues, using a Kalman filter, insert a jump at 9/11, which does leads to a far better fit (Beck, Jackman, and Rosenthal 2006). Inserting the jump at 9/11, though, assumes the answer to the question we are asking. While admitting 9/11 as a break seems sensical from even a casual glance at the data, an equally strong case could be made for a wide variety of events, such as the invasion of Iraq or Afghanistan, the events surrounding Hurricane Katrina, and so on. We simply do not know when to stop adding jumps.

This brought to the fore the two questions central to our method. Given data with a time component (or any natural ordering), we develop a method that identifies both the location and the number of jumps, while fitting a smooth curve elsewhere. The method fits a smoothing spline to the data, while using a binary segmentation algorithm to sequentially add "jumps" to the spline's unpenalized space. We develop a modified Bayesian Information Criterion (BIC) statistic as a stopping rule.

We use the method here on two different sets of observed data. The first is George W. Bush's approval ratings across multiple pollsters between January 25, 2001, and October 24, 2007. Our algorithm picks out two critical events in Bush's term plus or minus a day or two: September 11, 2001, and the invasion of Iraq. We also analyze the median of the first dimension DW-NOMINATE, both by Congress and by party. With the DW-NOMINATE medians, we detect jumps in the Congressional median between 1910–1912 and 1874–1876, most likely

corresponding to the rebellion against Joseph Cannon and the end of Reconstruction. We find a jump between 1930 and 1932 for the Democratic party, corresponding to the beginning of the New Deal. We find a jump for the non-Democratic party in 1818–1820, which is most likely a false positive.

The method provides several advances over existing methods. Although there are several standard methods for finding a structural break in a parameter (as in Calderia and Zorn 1998, e.g.), ranging from the simple Chow test to Bayesian methods that estimate the location of the change point (Western and Kleykamp 2004), most do not offer a technique for finding both the location and the *number* of breaks. Recent work by Arthur Spirling has addressed this issue explicitly, through the use of reversible jump Markov chain Monte Carlo, in finding "turning points" in civilian casualties through the Iraq war (Green 1995; Spirling 2007b). In contrast to Spirling, we embed our method within a nonparametric framework through the use of smoothing splines. The method employed by Spirling allows for identification of breaks in a given parameter in a structural model, whereas we search for a series break in the mean function itself, rather than any particular parameter. We are able to avoid heavy parametric assumptions about either the systematic or the random component of the model, which allows extension to a broad array of spline models and semiparametric models (Gu 2002).

A rather straightforward change in loss function could easily allow for modeling limited dependent variables, similar to methods that search for breaks in parameters in generalized linear models (Spirling 2007a). This would allow for the identification of change points within limited dependent variables. Finally, to ease interpretation, our method returns a serial order of breaks and a corresponding modified BIC statistic. This allows the researcher to present the jumps in terms of importance, as well as flexibility in the stopping decision.

2 Methods

2.1 *A Brief Review of Smoothing Techniques*

Our interest with the presidential polling data is in summarizing its progression over time in a clear way while accounting for noisy measurements taken at irregular times. Any summarization is a trade-off between an estimate that is too variable and too complex to be readily interpretable and one that is too smooth and glosses over too much information. We search in between for a parsimonious, intelligible, de-noised estimate.

In the interest of producing an interpretable, data-driven estimate of the mean function, researchers familiar with polling data will be familiar with loess smoothing techniques for summarizing the data (Cleveland 1979; Beck and Jackman 1997). As a generalization of the weighted average, the loess balances the fit between a smooth curve and no smoothing by a bandwidth parameter.

Selection, or estimation, of this bandwidth parameter is a common characteristic of many smoothing methods. Recent work by Luke Keele (2006, 2008) has introduced a political science audience to the issues involved with semi- and nonparametric smoothing, and we refer the reader there for background.

We wish to highlight the use of smoothing splines for obtaining a good functional estimate. We rely here on the functions in the gss library in *R* (Gu 2002; R Development Core Team 2008). In this article, we use the restricted maximum likelihood (REML) algorithm to estimate the smoothing parameter throughout (Krivobokova and Kauermann 2007).

Given data (y_i, x_i), we assume that y_i is linear in a smooth function of x_i and a mean-zero error term that is independently and identically distributed with finite fourth moment. We also assume that the function is sufficiently "smooth," in that $\int \{f''\}^2 dt < \infty$.

$$y_i = f(x_i) + \varepsilon_i. \tag{1}$$

$$E(\varepsilon_i) = 0. \tag{2}$$

$$\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2. \tag{3}$$

The cubic smoothing spline fits a function that has two components. The first, unpenalized component is the linear trend to the data. The second, penalized component allows for a nonlinear fit. Controlled by a smoothing parameter, λ, the estimator is a compromise between the least squares line ($\lambda \to \infty$) and complete interpolation of the data ($\lambda = 0$). The cubic smoothing spline minimizes the penalized residual sum of squares:[1]

$$\hat{f}_{SS} = \min_f \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int \{f''\}^2 \, \mathrm{d}t. \tag{4}$$

The second component is the penalty on the ''nonlinear'' part. If f is the sum of a linear part and a smooth (C^∞) nonlinear part, for example, by admitting a Taylor expansion, then the linear part of the function will have second derivative zero and hence is not considered in the penalty. The smoothing parameter λ controls the trade-off between smoothness and the least squares regression line. The smoothing spline is widely applied, and software packages implementing cubic smoothing splines are common (Wahba 1990; Gu 2002).

2.2 *A Different Kind of Function*

These methods are applicable when we believe that the true function underlying our data is a smooth function measured with some noise. The 9/11 problem, though, highlighted an aspect of Bush's approval data that the smoothing spline could not handle. We are interested in a function that is everywhere smooth except for a small number of discontinuous jumps. We develop a method for identifying discontinuous shifts in the mean function, fitting a single smooth function but modeling breaks in the intercept. Discontinuities of this form represent a problem for smoothers. They can be viewed as a form of model misspecification or omitted variable bias. The smoother, when it smooths over the jumps, will leave a distinct residual pattern: residuals will be systematically above on one side of the jump and below on the other. Figure 2 illustrates the residual pattern around 9/11 for both the smoothing spline and a smoothing spline with a partial spline added at 9/11.

This misspecification creates a problem for selecting the smoothing parameter. Consider the jumps as sustained shifts in the mean from one part of the function to another. Residual variance estimates will be overstated, by adding part of the ''jump-to-jump'' variance to the true residual variance.

If we knew the location of jumps in a function, we might add them directly to the unpenalized space of the spline. This would augment the unpenalized space, resulting in a partial spline (Gu 2002). If $\varphi(x) = I\{x > \text{''9/11''}\}$, then the partial spline is the minimizer

$$\hat{y}_{PS} = \arg\min_{f,\beta} \sum_{i=1}^{n} (y_i - f(x_i) - \beta\phi(x_i))^2 + \lambda \int \{f''(t)\}^2 \, \mathrm{d}t. \tag{5}$$

---

[1] We formulate the problem in a least squares framework, but distributional assumptions could be added so as to characterize the data-generating process (see Gu 2002 for an extensive discussion with examples). This would result in a penalized likelihood rather than penalized regression.
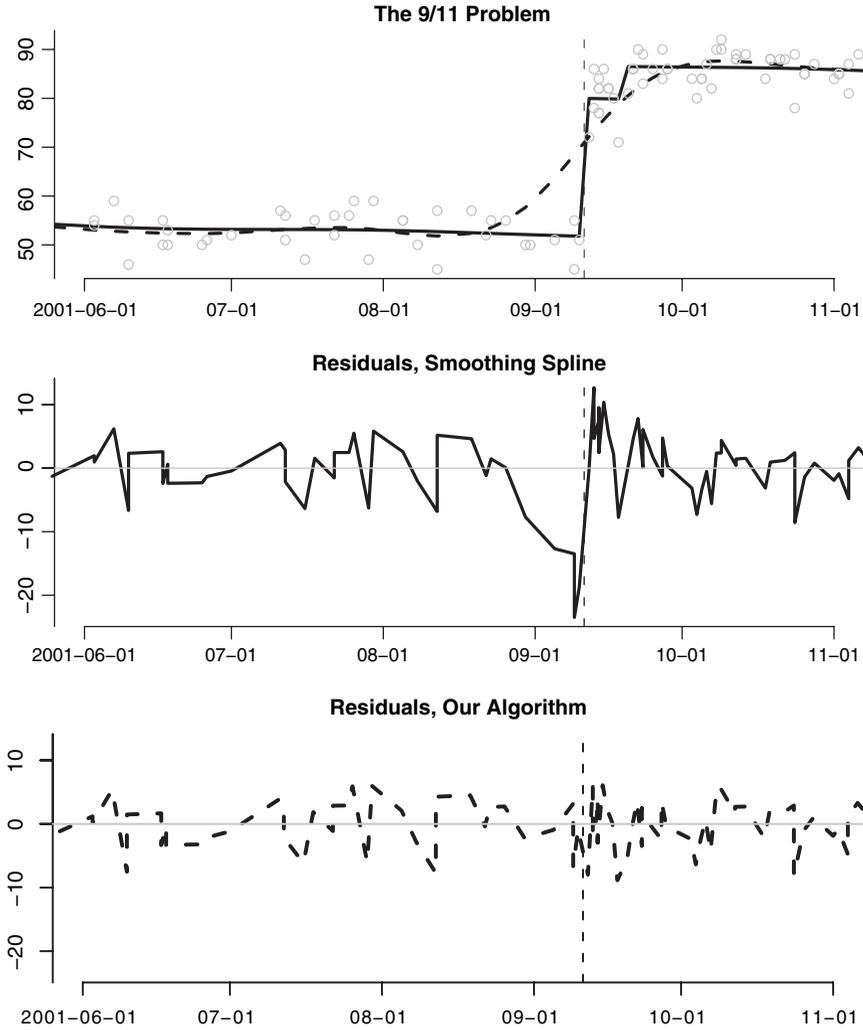
Fig. 2 The top shows the fit around 9/11; the bottom shows the residual pattern.

That is, known functions are not penalized and are accounted for in the fit directly. One might think of them as being initially subtracted out, and the remainder smoothed.

''Breaks'' in the residuals can be detected by eyeball and automatically. We suggest the application of binary segmentation procedures (Sen and Srivastava 1975), which look for breaks in a constant mean function.

Let $\hat{e}_1, \ldots, \hat{e}_n$ be the estimated residuals from the smoothing spline fit, after removing the linear trend. Define the partial sums $S_i = \hat{e}_1 + \ldots + \hat{e}_i$. Under the assumption of homogeneity and known variance, Sen and Srivastava derive the uniformly most powerful test for a break point. They show that the most likely break point can be found at the maximal t statistic $Z = \max_{1 \leq i \leq n} |Z_i|$ for

$$Z_i = \left(\frac{1}{i} + \frac{1}{n-i+1}\right)^{-1/2} \left(\frac{S_i}{i} - \frac{S_n - S_{i+1}}{n-i+1}\right). \tag{6}$$

We propose a slightly adjusted statistic. Denote $\hat{\sigma}_i$ as the estimated standard error of the first i residuals and $\hat{\sigma}_{i^-}$ as the standard error of the last i residuals. Let $Z^* = \max_{1 \leq i \leq n} |Z_i^*|$ for

$$Z_i^* = \left(\hat{\sigma}_i^2 + \hat{\sigma}_{(n-i+1)^-}^2 \right)^{-1/2} \left(\frac{S_i - \frac{S_i}{i}}{i} - \frac{S_n - S_{i+1} - \frac{S_n - S_{i+1}}{n-i+1}}{n-i+1} \right). \quad (7)$$

Z^* greatly outperformed Z in both the simulations and the observed data. We developed Z^* to account for the "messiness" in data normally encountered by political scientists. The residuals to the left and right of a given break point may be neither mean zero nor homoscedastic, and Z^* accommodates these possibilities, whereas Z does not. This better captures the nature of observational data, but our test is no longer uniformly most powerful, due to the unbalancedness of the design and ambiguity of degrees of freedom in the variance estimate.

Since we are not as interested in the inference problem, we need only assume that the data are reasonably symmetric about the mean function. The statistic Z^* serves to locate the most likely position of the discontinuity. It considers, at each possible break point, the extent to which the residual immediately to the left is above the mean to the left and the residual immediately to the right is above the mean to the right. The more these two terms differ, the more likely a discontinuous break occurs at each point. The weight term in front comes from the variance estimate, and experimentation revealed it effective in regulating erratic behavior at the boundaries.

This is a rough approximation. The pattern of residuals about the jump is not constant, but in the interest of finding a single break, binary segmentation is a reasonable, computationally simple approximation. As our examples below show, it is also quite feasible. We had experimented with modeling the strength of the break through a weighted average that aligns with residual patterns (e.g., wavelets, exponentially damped sine curves) as well as more exotic variable selection methods (Efron et al. 2004, e.g.), but the additional complexity of method added little to the performance of the algorithm. When choosing jumps, the simple statistic above performed as well as its more complex competitors.

2.3 Stopping Rules

Since this is an automated process, we require a stopping rule. We consider several here. The first is a modified BIC statistic, with $\hat{\sigma}_\varepsilon^2$ an estimate of residual variance, n the sample size, n^* the number of randomly selected knots,[2] and k the number of jumps:

$$\arg\min_k \frac{\hat{\lambda} \int \{\hat{f}''(t)\}^2 \, dt}{\hat{\sigma}_\varepsilon^2} + k \, \log(n) - \frac{k}{2}\log(n^*) + \frac{k}{2}\log(2\pi). \quad (8)$$

Appendix A contains a derivation of the traditional BIC statistic and our modification of it. Note how, in the event that the spline is fit to every observed data point ($n = n^*$), this statistic reduces to the traditional BIC statistic. The term with the $\log(2\pi)$ is asymptotically negligible since it does not grow in n or n^*, but simulations revealed that the term was helpful in reducing the false-positive rate for our smallest simulation ($n = 100$, $n^* = 30$). It did not make a large impact on the power of the larger simulations ($n = 250, 500$).

---

[2]Splines select a random subset of knots due to computational difficulties in inverting large matrices. Even with relatively few knots, estimates are quite stable and accurate.

The first term is increasing in k through both the numerator and the denominator. As more jumps are added, the spline has to bend less in order to accommodate the jumps, so the spline fit is more linear and the numerator increases. The penalty in the numerator is left unit-free by dividing through by the residual variance estimate. As well, the residual estimate decreases with the addition of more jumps, decreasing the denominator. The same logic motivating BIC then leads to a ''cost'' of $\log(n) - \frac{1}{2}\log(n^*)$ for the addition of each additional jump, in order to ensure convergence to the true model as $n \rightarrow \infty$. The spline penalty term serves the role of the log-likelihood, acting as a measure of the divergence between the ''true'' model and the fitted values.

As a means for balancing model fit and size, the BIC provides an estimate to the Bayes's factor. The BIC approximates the posterior probability that a given model is the true model, given a uniform prior over all candidate models. The statistic is not interpretable on its own but is useful in model selection. The most preferred model with the sequential smoothing spline is the one with the smallest modified BIC statistic as given by equation (8).

A second stopping rule comes from the logic of hypothesis testing. Although we recommend our modified BIC statistic, we present the following below. The reference distribution for the maximal Z is usually determined by permutation. This creates problems, though, because the reference distribution of each subsequent break must be conditioned on the selection of all previous breaks. The permutation method will increase in complexity rather quickly. As a simplification, we suggest considering the models as a sequential series of nested models. This suggests an approximate χ^2 statistic to consider among them:[3]

$$\frac{\hat{\lambda} \int \left\{ \hat{f}''\left(t\right)\right\}^2 \mathrm{d}t}{\hat{\sigma}_\epsilon^2} + c_{\alpha,k}^*,$$

where c^* is the critical value of the χ^2 statistic, with family-wise error rate α and number of breaks k:

$$c_{\alpha,\ k}^* = \chi^2_{1-\alpha/k,\ k}.$$

Using Bonferroni's adjustment of α, the term $1 - \alpha/k$ keeps the probability of making at least one type I error among all breaks below α. This approach has a noticeably lower threshold even at small sample sizes of 100 for the first several jumps; for this reason, we remain wary.

The final suggestion is using a combination of expert evaluation and common sense. As heuristics, we suggest a few options. First, if a selection may be a false positive, explore the selections afterward. If several selections after it are known jumps, the jump under question is more plausible. If there are no known jumps afterward, then its selection is less plausible. If the researcher has a specific prior distribution proportional to weights w_i over all possible jumps, the statistic $w_i\left|Z_i^*\right|$ may of course be used instead. Similarly, known breaks can be incorporated into the unpenalized space directly, and our method consequently implemented.

Second, we noticed that sometimes the BIC may be quite flat around the minimum. This may limit the discussion to the range of acceptable jumps, such as 3–5, even if the strict minimum occurs at 4. Similarly, the first few jumps result in a large decrease in the penalty

---

[3]Since we are not making any distributional assumptions about the error term, the χ^2 statistic is only an approximation.

term, but the effect decreases dramatically. Using either rule above, some jumps are clearly reasonable, some are questionable, and some are unreasonable. We suggest reporting either the BIC statistic or χ^2 p-values along with the data, and the researcher may decide to take the first of the "questionable" jumps as the stopping points. If computational power permits, resampling methods may be used, but we have been satisfied with the performance of choosing the modified BIC-minimizing number of partial splines. That is the rule we use throughout the remainder of the article.

Finally, with reasonably large data sets, the cubic splines are fit only using a random subset of the data as "knots." This introduces variance into the stopping rule statistic, so we recommend exploring either adding to the number of data points selected or a resampling method to help get a sense of the variance of the test statistics described here. With large data sets, of over 1000, increasing the number of knots can aid in detection, making the acceptance level of our BIC statistic less stringent. Increasing the number of knots, though, is not computationally prohibitive. Bush's approval data discussed below have an n of 1533, and ssanova selects only 52 knots. Computation took about 12 s on a Pentium D processor. Using our algorithm with 200 subsampled data points lengthened the time to only 2 min and 11 s.

These guidelines may sound arbitrary, but, as illustrated below, our method is never "too" wrong. It is conservative, in that it rarely overestimates the number of jumps, and even with reasonably small sample sizes and in the presence of autocorrelated noise, the method is powerful.

2.4 *Procedure Statement*

We suggest the following recursive procedure for finding these jumps in smooth functions:

1. Fit a cubic smoothing spline using the REML algorithm to estimate the smoothing parameter.
2. Remove the linear trend from the residuals.
3. Search the residuals for a break point, using the binary segmentation algorithm.
4. Add the corresponding break to the partial spline's unpenalized space.
5. Repeat from 1 until the BIC stopping rule ends the procedure.

This will result in breaks $\delta_1, \ldots, \delta_k$, and a mean function estimate of the form

$$\hat{y}_i = \hat{f}(x_i) + \sum_{j=1}^{k} \hat{\beta}_j I(x_i > \delta_j),$$

where $\hat{f}$ is a smooth spline estimate. The estimate $\hat{\beta}_j$ gives the estimated magnitude of the discontinuity and is reported when fitting partial splines. The standard errors on this coefficient will be wrong, though, since they do not account for the sequential nature of the selection and fitting. Were this magnitude of interest, and confidence intervals desired, we recommend a parametric bootstrap. Our algorithm sequentially augments the unpenalized space of the spline, fitting fixed jumps at specific points. Although estimating the number of jumps simultaneously may be preferred, and in more than one dimension is most certainly preferable, we are happy with the performance of the method in one dimension. We illustrate the procedure below on two observed data sets, Bush's approval and DW-NOMINATE scores, as well as through a series of simulations.

3   Case 1: George W. Bush's Approval

The data in this section are 1533 estimates of George W. Bush's approval, from 33 different houses, between January 25, 2001, and October 24, 2007. We observe polls on 1041 dates over the total 2463 dates. Although the polls are taken over multiple days, we consider the date range to be the end date on which the poll is conducted. The approval estimates range from 27% to 89%. The data are all publicly available.

Figure 3 shows the fit to each of the different methods, each in *R*. We use a moving average with a 2-week window, loess with spans .1 and .01 with command loess in library stats, the smoothing splines from command ssanova in library gss, a structural time series model using a Kalman filter from command StructTS in library stats, and our sequential segmentation method.

Results from each of the methods are shown in Fig. 3. Our sequential segmentation method outperforms, in terms of mean squared error (MSE), both the loess with span .1 and the smoothing splines. Our method performs worse, in terms of MSE, than the Kalman filter
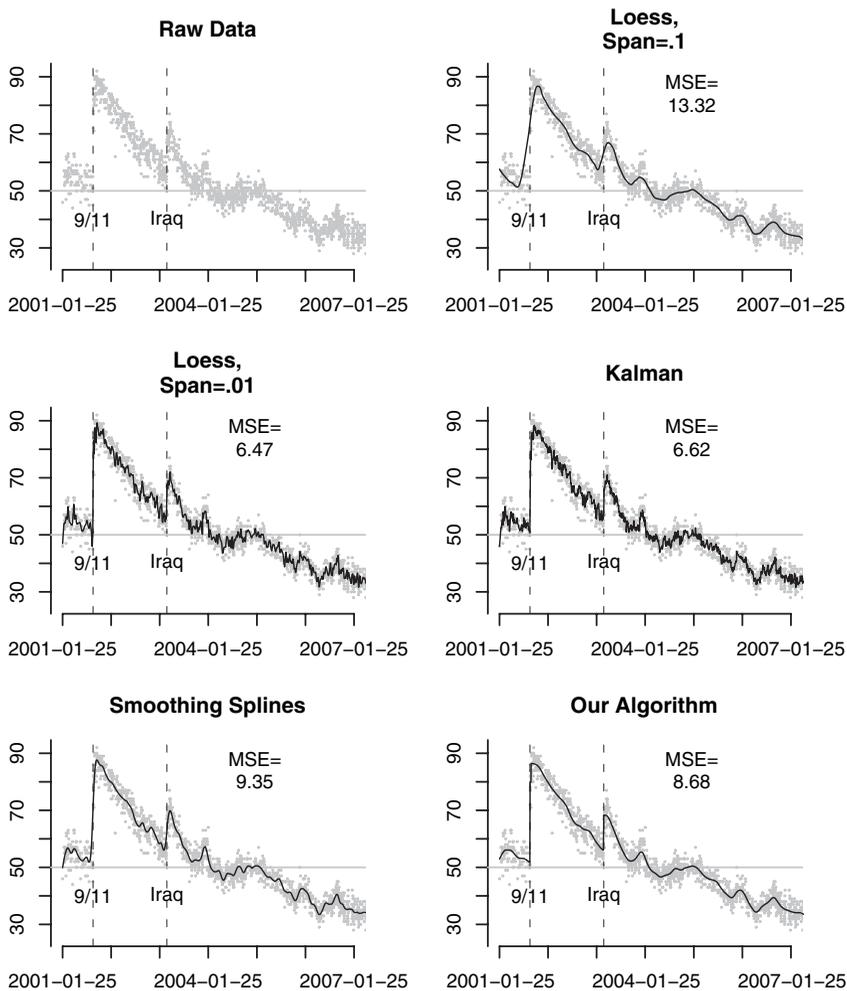

Fig. 3   A comparison of fits to the data among the different smoothing methods.
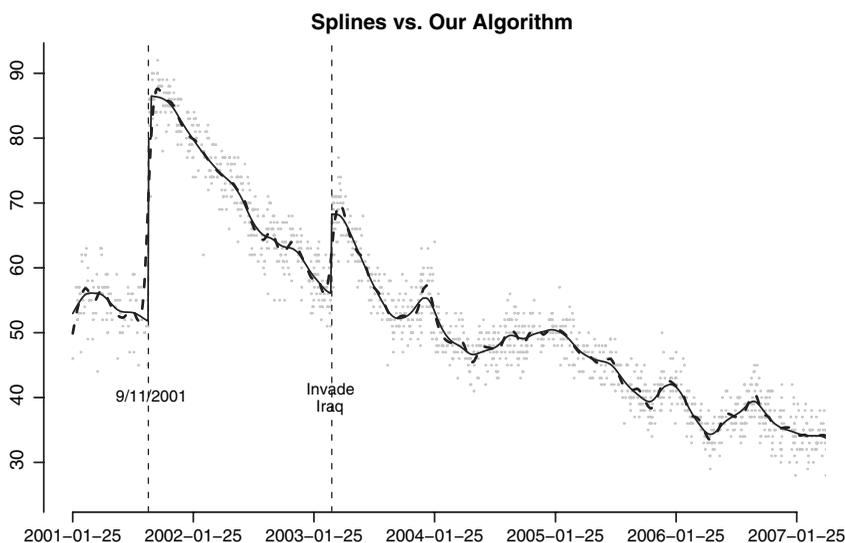
Splines vs. Our Algorithm

Fig. 4 Spline fit is dashed; our method fit is solid.

and the loess with span .01, but the improvement in MSE comes at the cost of a rather jagged fit. Figure 4 shows a larger graph, comparing splines to the sequential segmentation method. Notice how splines smooth over ''nonsmooth'' events. For example, the bump in Bush's ratings after 9/11 could not have started before the event; people had no expectations that the events would occur on a certain date and so they did not start revising their opinions of Bush upward before then. Note also how the ''rally around the flag'' effect, whereby a President's popularity increases upon starting a war, is immediate. During the run-up to the war, Bush's approval dropped regularly, but immediately upon the invasion, he got a bump up that then proceeded to erode. Our algorithm selected the ''rally around the flag'' effect as instantaneous, even when the invasion of Iraq was clearly expected.

Just as important, note how our method follows the spline method in areas after the events. From about mid-2004 on, our method follows the smoothing spline almost exactly. Our sequential segmentation method acts like a smoothing spline when appropriate.

The real payoff for the sequential segmentation method comes from looking at autocorrelation and the quantile-quantile (QQ) plots. Figure 5 shows the QQ plots, compared to a normal distribution, for the residuals using each of the six methods earlier. The other methods tend to do passably well in either having normally distributed errors (the second loess fit) or having low autocorrelation (the moving average, Kalman filter, and splines). Only our algorithm does well on both counts and noticeably better than the other methods.

Next, we consider the events selected as ''jumps'' in approval. The jumps and their corresponding BIC statistics are listed in Table 1. The trouble with evaluating whether these jumps are valid is two-fold. First, we have no idea what the ''true'' curve is or even if the concept of such a curve is theoretically coherent. We conceive of this curve, instead, as the consensus of approval ratings among pollsters. Looking at the dates chosen and seeing if anything happened involving approval considerations of Bush lead to a problem, though: given any day, you can find *something* of import that might affect approval toward
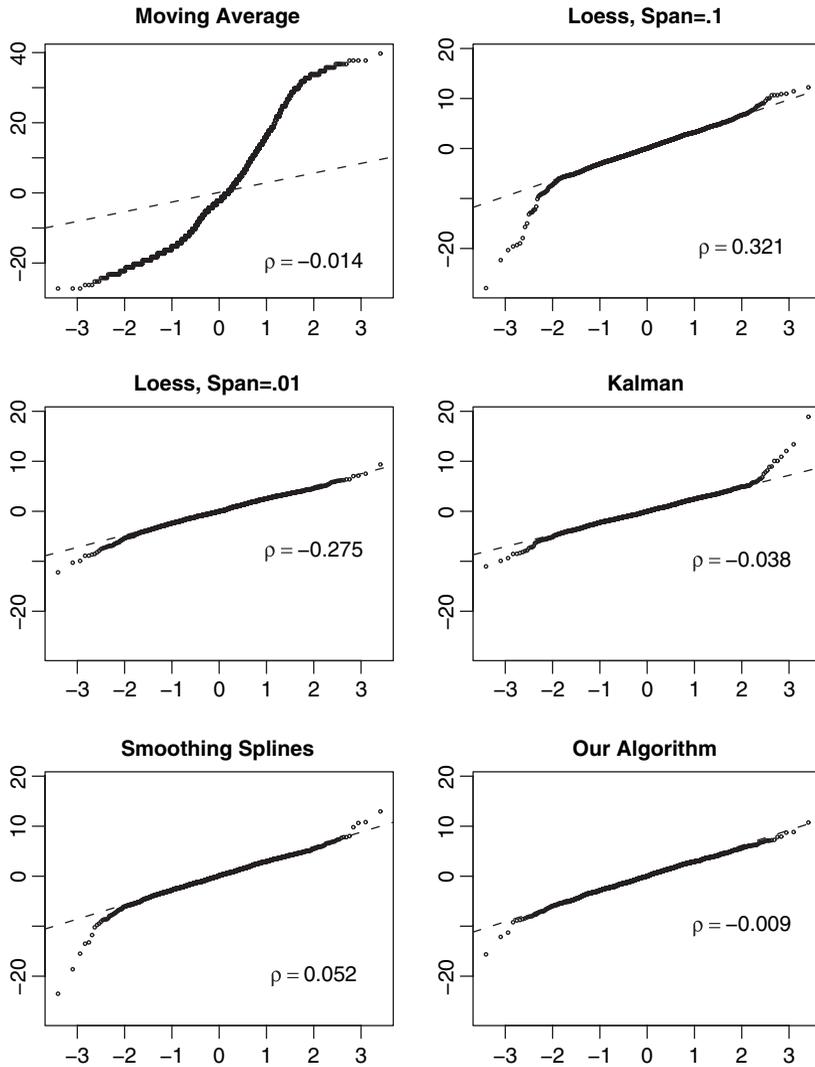
Fig. 5 QQ plots, for a normal distribution, for each of the smoothing methods. "ρ" gives the correlation in the residuals with lag 1. For ease of interpretation, the last five plots are on the same *y* axis.

Bush. Unfortunately, there is no objective listing, serial or otherwise, to which we can compare these results.

That said, we are confident in the dates selected. We are counting the date of each poll as the last day it was administered, so we look for critical events a few days before the breaks selected by our method. The two dates selected, September 12, 2001, and March 20, 2003, clearly correspond with major events that would plausibly affect approval scores in the Bush presidency (9/11, the invasion of Iraq on 3/17). Since we are taking the date of the poll as the final day of its administration, the method selects dates a few days past the event of interest.

Table 1 The first five events in Bush's term selected by the sequential segmentation spline

Selected date	Event	Actual date	BIC
Spline only	XX	XX	78.05
September 12, 2001	9/11	September 11, 2001	68.62
March 20, 2003	Invade Iraq	March 20, 2003	60.78
September 21, 2001	Bush's post 9/11 address to Congress	September 21, 2001	63.67
January 22, 2004	Bush's State of the Union	January 22, 2004	72.17
October 04, 2006	??	??	78.12

Note. The BIC criterion selects the first two events. We include the next three events to illustrate behavior of the BIC statistic.

The primary substantive conclusion is that presidential approval and quite likely other public mood trends are not smooth. Whereas recent theoretical work on macropolitical evolution views presidential approval as a smooth, AR(1) time series in Gallup approval (Erikson, MacKuen, and Stimson 2002), we find that future empirical work should search for and model multiple structural breaks.

4   Case 2: Congressional Ideology

The data in this section are three time series: the median estimates of the first dimension of DW-NOMINATE ideology scores by Congress and party. We consider two parties, the Democrats and the non-Democrats, a pooling of Republicans and Whigs for the sake of this article. We have data on each Congress from 1788 to 2004,[4] the non-Democratic Party from 1800 to 2004, and the Democratic Party from 1794 to 2004. Higher scores are commonly interpreted as being more ''conservative'' on economic issues, whereas lower scores mean that the Congressional median is more ''liberal'' on economic issues (Poole and Rosenthal 1997). The data are publicly available.

Unlike Bush's approval data, the median DW-NOMINATE data are evenly spaced through time, although with only one data point per Congress. Relative to Bush's approval, this is a relatively small-*n* test of the method, with only 109 data points, versus 1533 above.

A second key difference is that, in these data, no apparent discontinuities appear. The raw data for the entire Congress can be found in the upper left-hand corner of Fig. 6, with fits from various methods in the subsequent boxes. We notice nothing in these data comparable to the ''9/11'' problem above.

Despite the lack of visually obvious jumps, a long literature has debated as to whether some elections may be ''critical'' (Key 1955, 1959; Mayhew 2002). Most Congressional elections are ''secular,'' reflecting slow gradual change, whereas some scholars consider a select few as ''critical,'' in that they result in rapid, sustained shifts in partisan composition of the public. This dynamic, as expressed through Congress members, captures the nature of the ''smooth + jump''–type function that we are estimating here.

As suggested by the realignment literature, we rely rather heavily on our model's assumptions: that median NOMINATE score is the sum of a smooth and discontinuous

---

[4]For clarity, we refer to each Congress by the year in which it was elected rather than its number, that is, ''2006'' rather than ''110th.''
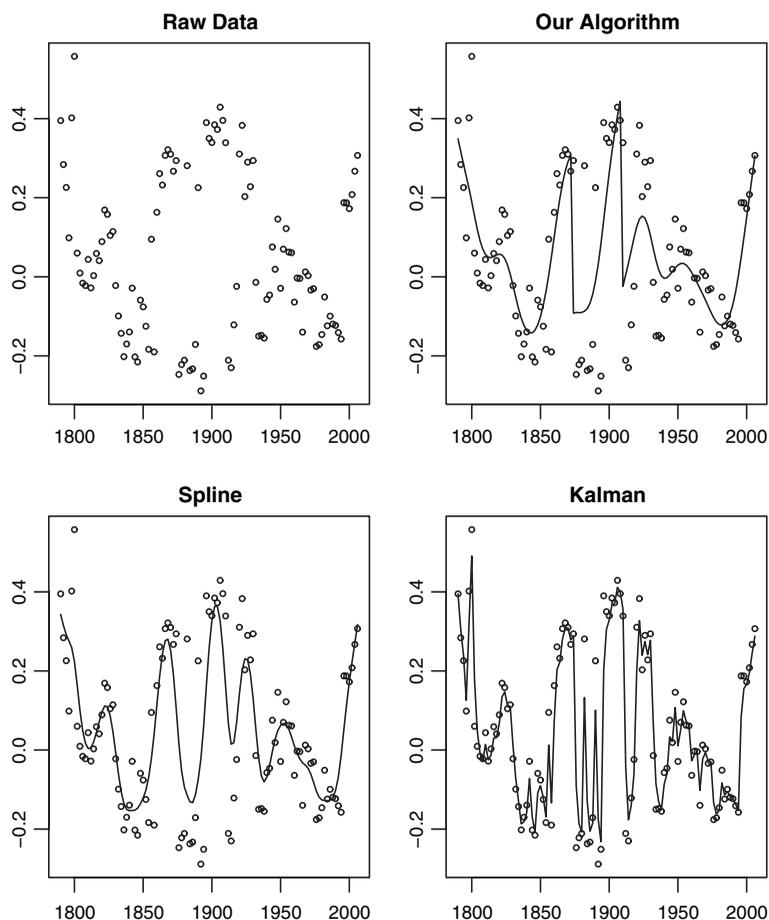
Fig. 6 A comparison of fits to the median DW-NOMINATE score by Congress among three different smoothing methods.

part. Comparing the fit of our method to the spline and time series method highlights how statistical findings are dependent on the assumptions under the statistical model. A researcher searching for smooth cycles in Congressional history may prefer either of the bottom two graphs. A researcher searching for a smooth curve with a few jumps would prefer our model. Our algorithm, though, is both consistent with the data and reduces to a spline in the limiting case.

Most of the discovered jumps, as shown in Table 2, are easily attributable to commonly accepted shifts in Congressional behavior. The two shifts selected for the entire Congress via BIC are 1910–1912 and 1874–1876. The first can be attributed to the rebellion against Speaker Joseph Cannon, which shifted power from the Speaker to the committee chairs. The second can be attributed to the end of Reconstruction. The next few dates, 1932, 1896, and 1860, though not selected by BIC, have long been considered the canonical dates for "realignment" (Mayhew 2002). The 1930–1932 break for Democratic median votes is clearly attributable to the New Deal. A single shift, in 1818–1820, was selected for the non-Democratic party. We worry this may be a false positive and may be a random artifact of pooling all non-Democratic parties.

Table 2 The first few events in Congressional history selected by the sequential segmentation spline

	Date	Event	BIC
Entire congress	None	Spline only	23.79
	1910–1912*	Joseph Cannon Rebellion	22.31
	1874–1876*	End of Reconstruction	21.58
	1932–1934	New Deal Realignment	23.11
	1896–1898	1896 Realignment	26.87
	1860–1862	Civil War	29.74
Non-Democratic party	None	Spline only	20.90
	1818–1820*	Possible false positive?	19.67
	1852–1854	??	22.79
	1856–1858	??	25.05
Democratic party	None	Spline only	14.92
	1930–1932*	New Deal Realignment	14.59
	1892–1894	Populist Realignment	17.53
	1954–1956	??	19.87

Note. Dates marked with an asterisk are selected by our BIC criterion.

The algorithm performed well, even in this noisy, imprecise setting. Plotting the data revealed no apparent, obvious jumps, and condensing the NOMINATE scores down to 109 points is a gross oversimplification of Congressional behavior and evolution. Given all this, our method yielded reasonable results, selecting only one false positive and a series of dates otherwise that correspond with well-known shifts in Congressional makeup and behavior.

5 Simulations

We conduct 24 separate simulations in order to evaluate our method. We assume two different systematic components, where f is a Bessel function of the second type and x in the interval [0, 1000]. We ran each simulation 1000 times each, comparing our algorithm to both smoothing splines (function ssanova in R library gss) and a Kalman filter (function StructTS in R library stats). The characteristics of the simulations are below.

Model specifications:

$$\text{sim}_{\text{jump}}(x_i) = -2f(x_i/100) + 8I(x_i > 200) - 4I(x_i > 500) + 2I(x_i > 800) + u_i,$$
$$\text{sim}_{\text{nojump}}(x_i) = -2f(x_i/100) + u_i.$$

Variance specifications:
Gaussian noise : $\text{var}(u_i) \in \{1, 4\}$; $\text{cor}(u_i, u_{i-1}) = 0$,
AR(1) noise : $\text{var}(u_i) \in \{1, 4\}$; $\text{cor}(u_i, u_{i-1}) = .4$.
Sample sizes used:
$n \in \{100, 200, 500\}$.

Figure 7 contains an example of the simulation with jumps as well as one of our fits to it. The first jump is significant at the 5% level in all situations. The second jump is significant only in our less noisy simulations, and the third jump is never significant at 5%.

We vary the simulations by sample size, error variance, and variance structure. Table 3 contains a list of each of the simulations we conduct for both sim_{jump} and $\text{sim}_{\text{nojump}}$, as well as the percent of times each number of jumps was selected by our algorithm. The simulations all contain either independent, Gaussian noise or AR(1) noise, with correlation .4.
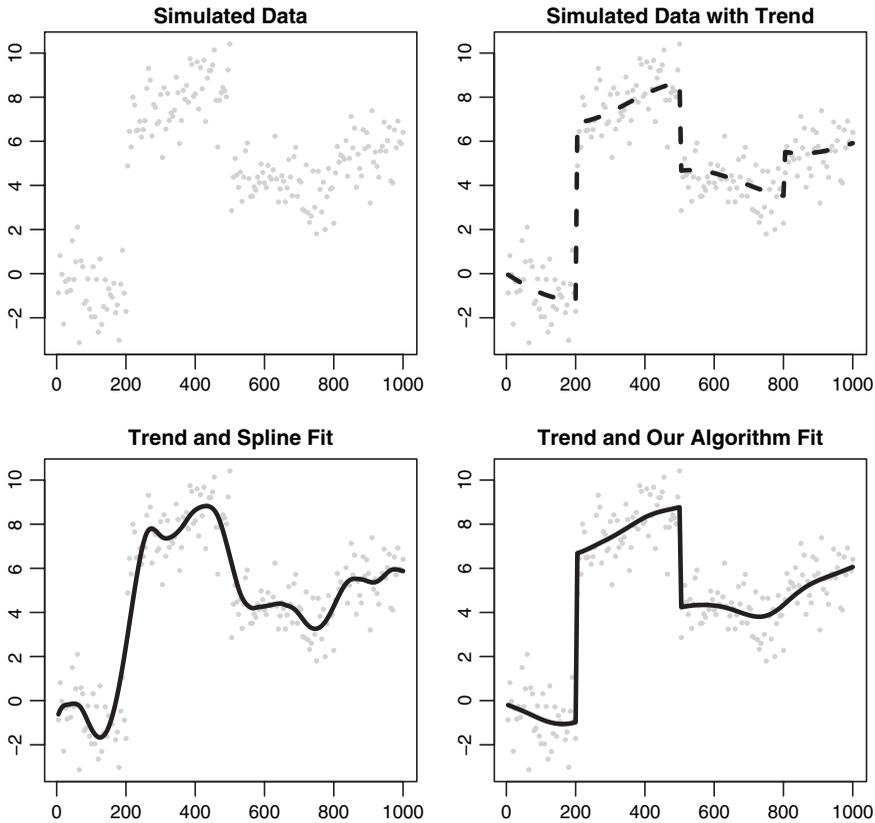
Fig. 7 An example from a run of the simulation, with $n = 200$, Gaussian noise, and $\text{var}(u_i) = 1$.

Our simulations demonstrate four desirable aspects of our algorithm. First, as shown in the bottom half of Table 3, the algorithm has a low false-positive rate when there are no jumps in the model. When there are no jumps, the algorithm will reduce to the spline, with an acceptable false-positive rate. For small sample sizes, and in our less noisy scenarios, the false-positive rate is approximately 10%, but the false-positive rate drops dramatically as the sample size gets to 500. False-positives rates are acceptably low in the other simulations, even for $n = 100$. The method is robust against error misspecification, and the false-positive rate drops to zero as the sample size increases. When there are no jumps in the true function, our method uncovers false positives at an acceptable rate.

Second, the method is powerful. Table 4 shows the percent of the time each of the three breaks was selected. The largest break, at $x_i = 200$, is selected the most. By $n = 500$, our algorithm selects the first break between 89% and 99% of the time. The next two breaks are selected less than the first break, as expected, since they are smaller in magnitude. The discovery rate of each of the smaller breaks, though, is clearly increasing in n. All these results are robust to error misspecification. Just as importantly, the false-positive rate is acceptable, always below 5.5 and well below 1% in 5 of the 12 specifications with jumps.

Third, the algorithm provides, on average, gains in squared error, and at best, the algorithm results in only modest losses versus the smoothing splines and time series method, as illustrated in Table 5. We calculated squared error as the sum of the squared difference between the fitted and true values. As shown in the top half of Table 5, our

Table 3 Percent of times each number of breaks was chosen, by simulation

	n	$var(u_i)$	0	1	2	3	4	5	6
Simulations with jumps									
Gaussian noise	100	1	63.2	27.2	6.6	2.4	0.6	0.0	0.0
	200	1	7.0	56.8	36.2	0.0	0.0	0.0	0.0
	500	1	0.8	7.5	89.2	2.5	0.0	0.0	0.0
	100	4	68.9	18.6	12.5	0.0	0.0	0.0	0.0
	200	4	36.3	39.2	24.5	0.0	0.0	0.0	0.0
	500	4	7.2	74.5	18.3	0.0	0.0	0.0	0.0
AR(1) noise	100	1	67.6	23.2	6.5	2.1	0.6	0.0	0.0
	200	1	19.6	32.7	39.6	7.3	0.8	0.0	0.0
	500	1	8.3	5.5	45.4	39.0	1.8	0.0	0.0
	100	4	34.1	41.9	20.6	2.9	0.5	0.0	0.0
	200	4	22.8	48.8	25.5	2.5	0.4	0.0	0.0
	500	4	10.3	33.3	51.1	4.7	0.6	0.0	0.0
Simulations without jumps									
Gaussian noise	100	1	89.2	10.8	0.0	0.0	0.0	0.0	0.0
	200	1	88.9	8.8	1.3	0.6	0.3	0.1	0.0
	500	1	99.4	0.2	0.1	0.0	0.3	0.0	0.0
	100	4	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	200	4	96.9	1.8	0.8	0.2	0.3	0.0	0.0
	500	4	99.9	0.1	0.0	0.0	0.0	0.0	0.0
AR(1) noise	100	1	92.8	6.5	0.7	0.0	0.0	0.0	0.0
	200	1	98.2	1.6	0.2	0.0	0.0	0.0	0.0
	500	1	99.7	0.3	0.0	0.0	0.0	0.0	0.0
	100	4	91.6	7.8	0.5	0.1	0.0	0.0	0.0
	200	4	92.3	7.4	0.3	0.0	0.0	0.0	0.0
	500	4	96.8	3.2	0.0	0.0	0.0	0.0	0.0

Note. There are three jumps total, two of which are easily discernible and one which is not.

method performs well relative to the other methods. The Kalman filter proves marginally better, about 4%, in 2 of the 12 simulations with jumps. Improvements in squared error over the Kalman filter and splines in the remaining simulations range between 1% and 55%. Of the 12 simulations with jumps, our algorithm outperforms the Kalman filter by more than 20% 9 of the 12 times. Our algorithm similarly outperforms the splines by more than 20% 6 of the 12 times.

The bottom half of Table 5 illustrates the results when the systematic component does not contain jumps. In every instance, our algorithm outperforms the Kalman filter, ranging between 6% and 83% gains in squared error loss. Our method performs comparably to the smoothing spline. The only exception is in the smallest, least noisy simulation, where splines outperform our algorithm by 19%. Of the remaining 11 simulations, the spline never provides an improvement of more than 6%, and in 6 of the 12 simulations, our algorithm differs from the smoothing spline by less than 1%.

Finally, the method is robust with respect to correlated errors. While our algorithm is less powerful in the presence of AR(1) noise around the discontinuous, smooth function, by modest sample sizes of 200, the algorithm performs well. Even with AR(1) noise, our method maintains a small false-positive rate. In the presence of breaks and AR(1) noise, our algorithm outperforms splines in each case, and in the absence of breaks, splines never provide more than a 6% advantage in squared error over our algorithm. In 2 of the

Table 4 Percent of the time each break was identified and the false positive rate, by simulation

	n	$var(u_i)$	ID 200	ID 500	ID 800	False-positive rate
Gaussian noise	100	1	84.4	14.8	0.0	0.6
	200	1	93.0	36.2	0.0	0.0
	500	1	99.2	91.6	2.4	0.2
	100	4	30.7	11.4	0.0	1.5
	200	4	63.6	22.1	0.0	2.5
	500	4	92.8	17.8	0.0	0.5
AR(1) noise	100	1	34.0	7.9	1.1	1.9
	200	1	83.0	47.7	5.8	0.5
	500	1	93.5	86.2	40.1	0.7
	100	4	67.3	20.4	0.8	5.3
	200	4	77.4	27.3	0.9	3.3
	500	4	89.8	55.8	4.1	2.3

12 simulations with AR(1) noise, our algorithm performs comparably to the Kalman filter; in the remaining 10, it provides substantial improvement, ranging between 25% and 86%.

6 Extensions of Method

Our method admits several interesting and useful extensions. First, we have presented it so far as a modeling, rather than inferential, tool, although resampling methods could provide confidence intervals. Since the search for breaks is sequential, the nature of the distribution under the null hypothesis is not clear; each subsequent break is conditioned on the occurrence of the earlier breaks. Entering unpenalized covariates of interest into the nonpenalized space also allows for inference in a semiparametric setting.

Second, our method can extend to a broad array of spline models.[5] These include extensions to higher dimensions, through thin-plate splines (Pearce and Wand 2006). Our method can help find jumps in geographic data or any situation in which the context dictates estimating a functional form that is smooth with breaks. Since the spline is a limiting case, and our cases above show that the method can "act like a spline" when it needs to, we hope to generalize the algorithm to where it can be useful across a broad array of data sets and questions.

Third, we have so far characterized our jump covariates as a series of indicator functions. The result is a form of a sequential cumulative sum tests, with a spline fit in between each identification of a jump. The jumps, though, could instead contain covariates that are of interest to the researcher. For example, using our Bush data, we could look for jumps in approval as a function of economic news, Congressional approval, etc. This would require simply ordering the outcome by a different covariate than time and using our sequential segmentation spline along this dimension.

Finally, we have developed the method so far with few constraining distributional assumptions about either the systematic component or the error. Stronger assumptions can be made, as necessary, in order to characterize a likelihood that could handle choice models, count models, and other limited dependent variable models.

---

[5]Specifically, any function that lives in a Hilbert space and has a reproducing kernel. See Wahba (1990) for extensions.

Table 5 MSE across simulations

	n	$var(u_i)$	MSE, Kalman	MSE, spline	MSE, our algorithm	% Improvement, versus Kalman	% Improvement, versus spline
Simulations with jumps							
Gaussian noise	100	1	0.609	0.696	0.632	−3.83	9.20
	200	1	0.481	0.513	0.380	20.89	25.86
	500	1	0.338	0.355	0.158	53.16	55.42
	100	4	1.549	1.234	1.192	23.05	3.44
	200	4	1.156	0.879	0.718	37.85	18.28
	500	4	0.763	0.566	0.347	54.59	38.73
AR(1) noise	100	1	3.641	3.648	3.601	1.10	1.30
	200	1	0.356	0.453	0.368	−3.38	18.85
	500	1	0.306	0.330	0.166	45.85	49.79
	100	4	1.316	1.036	0.937	28.82	9.57
	200	4	1.138	0.753	0.590	48.20	21.74
	500	4	0.942	0.506	0.287	69.56	43.27
Simulations without jumps							
Gaussian noise	100	1	0.168	0.069	0.082	51.31	−18.43
	200	1	3.631	3.434	3.405	6.21	0.84
	500	1	3.527	2.978	2.970	15.80	0.27
	100	4	0.444	0.260	0.261	41.36	−0.05
	200	4	13.867	12.369	12.336	11.04	0.27
	500	4	13.897	11.080	11.084	20.24	−0.04
AR(1) noise	100	1	0.215	0.076	0.080	62.70	−5.62
	200	1	0.174	0.044	0.045	74.17	−2.44
	500	1	0.137	0.019	0.019	86.03	0.18
	100	4	0.723	0.265	0.273	62.31	−3.01
	200	4	0.600	0.145	0.151	74.81	−4.21
	500	4	0.494	0.065	0.067	86.38	−3.45

Note. The last two columns show the average percent improvement of our algorithm over the other two methods.

7 Conclusion

This project began with the goal of estimating house-level effects on presidential approval. In estimating these effects, it quickly became apparent that extant smoothing methods missed the underlying dynamic of presidential approval. In fact, within much political data of interest, ''smoothing'' appears to be presumptuous. Many processes are a mixture of both slow-moving change and immediate jumps due to rapid shocks. Smoothers miss the nature of the underlying function, whereas structural break tests offer no stopping criterion. This led us to consider first, how to think about these processes and, second, how to find the breaks and when to stop doing so.

Our method has performed well both in simulations and on observed data. Within a dense presidential approval data set, our method discovered breaks that correspond with clearly identifiable shocks. Within a relatively sparse Congressional data set, it was equally successful at highlighting important dates. The simulations further reenforce our faith in the method.

The ''smooth + jump'' function we describe here could apply to a wide variety of political processes that face critical events. The algorithm allows a flexible means to model other social and physical processes of interest. Examples include change point

models or any model that must accommodate some smooth curve with the occasional persistent shock. In the face of a known exogenous shock, but uncertainty over the particular timing of the impact of the shock, our sequential segmentation spline will allow for estimation of the existence and most likely location of the discontinuity.

Testing for the existence of a structural break at a given point is straightforward. Here, we provide a method for solving a far harder problem: searching through all possible breaks, adding jumps sequentially, and stopping at a reasonable point. We plan to extend the method in the future to limited dependent variables and higher dimensional settings.

Appendix A: Derivation of the BIC and Our Modified BIC

This appendix provides a derivation of the modified BIC statistic used as our stopping rule. We follow the presentation by Adrian Raftery (1995), whereas a more technical discussion can be found elsewhere (Tierney and Kadane 1986).

Given observed data D, candidate models M_i, $i \hat{I} \{1, \ldots, m\}$, and associated parameters θ_i, the posterior can be written as

$$p(D|M_i) = \int p(D|\theta_i, M_i)p(\theta_i|M_i)\mathrm{d}\theta_i. \tag{A1}$$

Letting $g(\theta_i) = \log \{p(D|\theta)p(\theta)\}$, we now expand $g(\theta)$ around the posterior mode, $\tilde{\theta}_i$:

$$g(\theta) = g(\tilde{\theta}) + (\theta - \tilde{\theta})^T g'(\tilde{\theta}) + \frac{1}{2}(\theta - \tilde{\theta})^T g''(\tilde{\theta})(\theta - \tilde{\theta}) + o(||\theta - \tilde{\theta}||^2), \tag{A2}$$

g' and g'' denote the gradient vector and Hessian matrix of g with respect to θ. Since $\tilde{\theta}$ is the posterior mode, we know that the linear term above is zero, since $g'(\tilde{\theta}) = 0$. This leaves the approximation

$$g(\theta) \approx g(\tilde{\theta}) + \frac{1}{2}(\theta - \tilde{\theta})^T g''(\tilde{\theta})(\theta - \tilde{\theta}). \tag{A3}$$

Exponentiating both sides and noting that the integral is proportional to the integral of a multivariate normal give the result (MacKay 2003)

$$p(D) \approx \exp(g(\tilde{\theta})) \int \exp\left(\frac{1}{2} \cdot (\theta - \tilde{\theta})^T g''(\tilde{\theta})(\theta - \tilde{\theta})\right) \mathrm{d}\theta \tag{A4}$$

$$= \exp\left(g(\tilde{\theta})\right)(2\pi)^{\frac{d}{2}}|A|^{-\frac{1}{2}}. \tag{A5}$$

Assume that A approaches infinity as some nA_0, with A_0 the asymptotic information matrix. Then $\log|A|$ approaches infinity as $\log|nA_0| = \log(n^d|A_0|)$. Substituting this into the equation above and dropping all terms that are not changing in n or i leave

$$\lim_{n \to \infty} \log p(D) \approx \lim_{n \to \infty} \log p(D|\hat{\theta}_i) - \frac{1}{2}\log|A| \tag{A6}$$

$$= \lim_{n \to \infty} \log p(D|\hat{\theta}_i) - \frac{1}{2}\log(n^d|A_0|) \tag{A7}$$

$$= \log p(D|\hat{\theta}_i) - \frac{d}{2}\log(n). \tag{A8}$$

The first term on the right-hand side above is the log-likelihood, whereas the second is a measure of the dimensionality of the model. BIC is an asymptotic result, balancing the trade-off between model likelihood and dimensionality, with the model corresponding to the highest BIC generally selected. Often, it is written as $-2 \cdot$log-likelihood $+ d \cdot \log(n)$, in which case the model with the smallest BIC would be selected. The BIC provides an estimate of the posterior probability given a uniform distribution across all candidate models. If the correct model is a candidate, it will be chosen with probability one as the sample size approaches infinity.

We modify the BIC to account for the fact that the generalized smoothing spline fits a spline to a random subset of knots, due to the difficulties involved in inverting large matrices. Assume, with sample size n, that the spline is fit to n^* knots, chosen at random. This leads to covariance matrices $|A|$ and $|A^*|$, each of which approaches the same $|A_0|$, as n and n^* increase.

Taking the approximations,

$$nA \approx n^*A^*, \tag{A9}$$

$$\frac{n}{n^*}A \approx A^*, \tag{A10}$$

$$\log|A^*| \approx \log\left\{ \left(\frac{n}{n^*}\right)^d |A| \right\}. \tag{A11}$$

Given that A grows as nA_0, this leaves

$$\log|A^*| \approx \log\left\{ \left(\frac{n}{n^*}\right)^d |nA_0| \right\} \tag{A12}$$

$$= \log\left\{ \left(\frac{n^2}{n^*}\right)^d |A_0| \right\} \tag{A13}$$

$$= 2d\log(n) - d\log(n^*) + \log|A_0|. \tag{A14}$$

Substituting equation (A14) into equation (A6) leaves our modified BIC statistic, where the first term in the sum is the divergence of the sampled data, y^*, which is sampled from the complete set of observed outcomes, y:

$$\text{BIC}(n, n^*, i) = \log P(y^*|\hat{\theta}_i) - d\log(n) + \frac{d}{2}\log(n^*) + \frac{d}{2}\log(2\pi).$$

Rather than a likelihood, we use the penalty on the nonlinear part as the measure of divergence. The logic carries through identically upon noting the correspondence between the cubic smoothing splines used here and penalized regression (Pearce and Wand 2006). Note how, as n^* gets closer to n, our modified BIC approaches the actual BIC. Simulations with $n = 100$ and $n^* = 30$ indicated that the term with $\log(2\pi)$ was necessary to maintain a small false discovery rate in such a small sample. This asymptotically negligible term did not affect the power in the larger-n simulations, though.

We use this statistic as a stopping rule throughout our analysis and simulations.

References

Beck, Nathaniel, and Simon Jackman. 1997. Beyond linearity by default: generalized additive models. *American Journal of Political Science* 42:596–627.

Beck, Nathaniel, Simon Jackman, and Howard Rosenthal. 2006. *Presidential approval: the case of George W. Bush.* Working Paper.

Calderia, Gregory A., and Christopher J. W. Zorn. 1998. Of time and consensual norms in the supreme court. *American Journal of Political Science* 42:874–902.

Cleveland, William S. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74:829–36.

Efron, Bradley, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. 2004. Least angle regression. *Annals of Statistics* 32:407–99.

Erikson, Robert S., Michael MacKuen, and James A. Stimson. 2002. *The macro polity. Cambridge Studies in Public Opinion and Political Psychology.* Cambridge, United Kingdom: Cambridge University Press.

Green, P. J. 1995. Reversible jump MCMC computation and Bayesian model determination. *Biometrika* 82:711–32.

Gu, Chong. 2002. *Smoothing spline ANOVA models. Springer Series in Statistics.* New York: Springer.

Keele, Luke. 2006. *How to be smooth: automated smoothing in political science.* Unpublished Manuscript, Ohio State University.

———. 2008. *Semiparametric regression for the social sciences.* London: Wiley and Sons.

Key, V. O. 1955. A theory of critical elections. *The Journal of Politics* 17:3–18.

———. 1959. Secular realignment and the party system. *The Journal of Politics* 21:198–210.

Krivobokova, Tatyana, and G.öran Kauermann. 2007. A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association* 102:1328–37.

MacKay, David. 2003. *Information theory, inference, and learning algorithms.* Cambridge, United Kingdom: Cambridge University Press.

Mayhew, David. 2002. *Electoral realignments: a critique of an American genre.* New Haven, CT: Yale University Press.

Pearce, Nathan D., and Matthew P. Wand. 2006. Penalized splines and reproducing kernel methods. *The American Statistician* 60:233–40.

Pierson, Paul. 2004. *Politics in time: history, institutions, and social analysis.* Princeton, NJ: Princeton University Press.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: a political-economic history of roll call voting.* New York: Oxford University Press.

Raftery, Adrian. 1995. Bayesian model selection in social research. *Sociological Methodology* 25:111–63.

R Development Core Team. 2008. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Sen, Ashish, and Muni Srivastava. 1975. On tests for detecting the change in the mean. *The Annals of Statistics* 3:98–108.

Spirling, Arthur. 2007a. Bayesian approaches for limited dependent variable change point problems. *Political Analysis* 15:387–405.

———. 2007b. Turning points in the Iraq conflict: reversible jump Markov Chain Monte Carlo in political science. *The American Statistician* 61:1–6.

Tierney, Luke, and Joseph Kadane. 1986. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* 81:82–6.

Wahba, Grace. 1990. *Spline models for observation data.* Philadelphia: Society for Industrial and Applied Mathematics.

Western, Bruce, and Meredith Kleykamp. 2004. A Bayesian change point model for historical time series analysis. *Political Analysis* 12:354–74.