Bayesian Approaches for Limited Dependent Variable Change Point Problems

Arthur Spirling

Department of Political Science, University of Rochester, Rochester, NY 14620
e-mail: spln@mail.rochester.edu

Limited dependent variable (LDV) data are common in political science, and political methodologists have given much good advice on dealing with them. We review some methods for LDV "change point problems" and demonstrate the use of Bayesian approaches for count, binary, and duration-type data. Our applications are drawn from American politics, Comparative politics, and International Political Economy. We discuss the tradeoffs both philosophically and computationally. We conclude with possibilities for multiple change point work.

1 Introduction

Political scientists face substantively meaningful restrictions on the range of their chosen dependent variable in all areas of the discipline. This includes limited dependent variable (LDV) time series applications, where, for example, we record year-on-year Supreme Court dissent counts (e.g., Calderia and Zorn 1998) or (binary) occurrences of war between nations (e.g., Oneal and Russett 1997). We typically fit various generalized linear models (GLMs) to such data and there is much good advice available on the need for such approaches (e.g., King 1988), their problems (e.g., Beck, Katz, and Tucker 1998), and on interpreting their results (e.g., King, Tomz, and Wittenberg 2000). An overlooked aspect of LDV time series work relates to the proper analysis of *change points*: structural breaks in the direction or magnitude of quantitative relationships over time.

The (single) change point problem is rendered formally by supposing that Y_i represents our dependent variable of interest for which we have a total of T observations recorded through time $Y_1, Y_2, \ldots, Y_T$. The contention is that one data-generating process (DGP) characterizes the Y_i before a certain time and another after that point. That is, $Y_i \sim g(Y), i = 1, \ldots, k$, and $Y_i \sim h(Y), i = k+1, \ldots, T$, where $g(\cdot)$ and $h(\cdot)$ are known densities characterized by different parameters. Scholars are interested in k, the change point, which is *a priori* unknown and takes values in $\{1, 2, \ldots, T\}$. The characteristic parameters of $g(\cdot)$ and $h(\cdot)$ are also of concern since they demarcate the "effects" of the break; in a GLM context, they are a (linear) function of independent variables x with coefficient vector β.

We might thus talk, hypothetically, of a structural break for turnout in, say, 1980, with a coefficient sign switch on some predictor, like "race of respondent."

Though political scientists regularly conjecture the existence of such change points,[1] testing for them and their effects is rare. When testing *does* take place, scholars should be aware of the options available and the strengths and weaknesses of different approaches. These issues are *theoretical* in terms of the selection of a plausible model for the data and, indeed, an inference framework. They are also *computational* in terms of the practical ease of estimation, its sensitivity, robustness, and so on.

In this paper, we concentrate on *Bayesian* methods for LDV change point problems with an attendant discussion of the tradeoffs inherent in their application. This is in contrast to data problems for which ordinary least squares (OLS) (or normal likelihood models) are designed, for which Western and Kleykamp (2004) give a thorough discussion and demonstration of Bayesian and non-Bayesian options. Here, we deal with the philosophical appeal of three LDV estimators for count, binary, and duration problems derived from a logic initially expounded by Carlin, Gelfand, and Smith (1992). To wit, in Section 2 we review OLS and maximum likelihood (ML) approaches for LDV problems, of which perhaps the Chow test is best known. In Section 3, we introduce the Bayesian LDV estimators and to give readers a sense of their possible uses, we discuss applications to the Supreme Court, international currency exchange rate choice, and cabinet duration in France. We also consider some possible extensions to multiple change point situations. Section 4 concludes.

2 OLS and ML Approaches

Much like OLS itself, least squares tests for structural breaks are relatively well known in political methodology and are commonly available in statistical software packages like *Stata* or *R*. Typically, one of two scenarios is presented. First, suppose the researcher had a strong hypothesis about when the change took place and denote this point k (so, $k = 1980$ in our turnout example). The *Chow test* applies least squares estimation to the subsamples separately for observations 1 through k and then $k + 1$ through T, where T is the end of the series. An F test is then used to see if the imposed restriction of a break is significant. If there are insufficient observations for a given period (either 1 through k or $k + 1$ through T), life is marginally more difficult, though a variation on the Chow test is possible (Greene 2003, 130–1).

Suppose, alternatively, that the position of the change point is *not* known *a priori*. One option is the cumulated sum of residuals (*CUSUM*) test that is designed to deal with data where the break occurs gradually over time. The null hypothesis is that the coefficient vector for the regression is the same in every period, and the alternative is simply that the coefficient vector is not the same. This has the consequence that the CUSUM test may be used for an unknown change point problem though it lacks power relative to the Chow test.

There are more complex alternatives to the CUSUM, based on a generalized method of moments estimation procedure, and Greene (2003, 139–41) discusses them. One option is a ML estimate, which compares the log-likelihood function under the alternative

---

[1]For example, it is argued that the nature of legislative debate and process in France was radically transformed by a constitutional change (Huber 1996); that black turnout is dependent on particular political personalities competing in general elections (Tate 1991); that presidential approval fluctuates more than it used to (Wood 2000); that the granting of presidential clemency has been on a general decline since the 1950s, with a particularly noticeable break occurring with Reagan's incumbency (Ruckman 1997); that civilian attitudes toward war were profoundly different after the Great War (Mueller 1991); and that the end of the Cold War ushered in a new era of international relations.

hypothesis of model instability (structural break) with the null hypothesis of stability. Another approach calculates a Lagrange multiplier statistic and is slightly simpler to implement. In either case, the tests operate by taking some predefined sample period (say 0.15 through 0.85 T) and examining the null hypothesis of stability through that period versus the alternative of structural change at some *unknown* point during that same sample period. Notice that this is conceptually different to locating a particular k as a change point. A more agnostic alternative is to vary the sample period (ω) and calculate a Sup LR(ω) and Sup LM(ω) statistic which are simply the maxima of the likelihood ratio and Langrage multiplier statistics calculated over all the options of ω that the researcher chooses. However, the limiting (χ^2) distributional assumptions that hold for the original tests do not apply to the maxima.

To summarize, political scientists have access to several methods for locating change points, but they are typically designed (and optimal) for OLS cases, and/or require a sharp theory in terms of the break's location to be used. Given the simplicity with which such tests may be used, it is tempting to do so even when the dependent variable is limited in some way. Typically though, there is a more plausible model for the DGP that allows users to incorporate the restrictions on the dependent variable seen in practice (e.g., binary, nonnegative and discrete, nonnegative and continuous). There is also, arguably, a more philosophically pleasing way to deal with the change point problem and we expand on this now.

3 Bayesian Approaches

In the tests above, we obtain a *p* value (or equivalent) on the (binary) possibility of a break. Notice that testing several possible dates either requires employing the tests several times or limiting the dates under review. That is, the test does not treat the break as a *parameter* to be estimated *simultaneously* with the (vector of) regression coefficients. Bayesian methods can do exactly this though. This is arguably preferable on philosophical grounds too (see Gill 2002, 1–6, for a general discussion). To see how, notice that a Bayesian approach obtains a posterior distribution over (the support of) k and hence rather than using a *p* value to make the binary decision that a particular date "is or is not" the change point, we can talk of the *probability* of a break on a particular date.

Moreover, depending on how we operationalize our Bayesian model, we can ameliorate some small sample problems incumbent in ML operations. This is quite apart from the potential computational inconvenience that maximizing a likelihood in a noncontinuous space might involve, and the fact that the resultant models to be compared (with and without change point) are nonnested (see Clarke 2001, for an overview).

Of course, free lunches are rare, and the Bayesian approach is not without its costs: the routines we now discuss are certainly more computationally burdensome than, say, a Chow test, and convergence of the algorithms to their limiting distributions must be assessed to avoid potentially misleading inferences (see, e.g., Gill 2002, 389–410).

3.1 *LDV Models*

Recall Section 1: there, if $k = T$, then we have "no change" in the series (because, of course, every observation is drawn from the same DGP). Following Carlin, Gelfand, and Smith (1992), we can write the likelihood as the double product of the (independent and identically distributed [i.i.d.]) observations before the change point and after. Denote $\mathbf{Y} \equiv (Y_1, \ldots, Y_T)$ and then notice

$$\mathcal{L}(k \mid \mathbf{Y}) = \prod_{i=1}^{k} g(Y_i \mid \theta) \prod_{i=k+1}^{T} h(Y_i \mid \eta), \qquad (1)$$

where θ (η) is a vector of parameters from the original (changed) DGP. By maximizing this likelihood we could obtain k directly. But a Bayesian approach places a prior distribution of $\tau(k)$ on $\{1, 2, \ldots, T\}$ as well as priors $p(\eta)$ and $p(\theta)$ on the other two parameters. We update our priors by the data we have (our observations, Y_i) to obtain a joint posterior distribution of k, θ, $\eta | \mathbf{Y}$:

$$\Pr(k, \theta, \eta \mid \mathbf{Y}) = \frac{\tau(k)p(\theta)p(\eta)\mathcal{L}(k|\mathbf{Y})}{\sum_{i=1}^{T} \tau(k)p(\theta)p(\eta)\mathcal{L}(k|\mathbf{Y})}. \qquad (2)$$

Dropping the normalizing constant, we have the more standard

$$\Pr(k, \theta, \eta \mid \mathbf{Y}) \propto \tau(k)p(\theta)p(\eta)\mathcal{L}(k|\mathbf{Y}), \qquad (3)$$

from which we will want to calculate various statistics, such as the median value of the change point. That will involve obtaining $\Pr(k | \theta, \eta)$. We will typically also be interested in $\Pr(\theta | \eta, k)$ and $\Pr(\eta | \theta, k)$ which will tell us about the direction and magnitude of the change in terms of the GLM coefficient vector. Any of these marginals will require integration of the joint posterior, which may be mathematically difficult, if not impossible. Markov chain Monte Carlo techniques will resolve this *impasse* through, in particular, the Gibbs sampler (see Casella and George (1992) or Jackman (2000) for an explication and discussion with examples). The Gibbs sampler requires that we are able to write down the relevant marginal posteriors from which sampling should take place even if we cannot do this, software packages like BUGS will implement a Metropolis-Hastings sampler and all that is required is the appropriate link function (Poisson, logit, and so on) and specification of priors. Though it may be computationally expedient to specify a conjugate prior for the coefficient vectors—that is a prior in the same family as the densities $g(\cdot)$ and $h(\cdot)$—it is not necessary to do so with *BUGS* which automatically adjusts its sampling scheme to accommodate such situations.[2]

There are two other modeling decisions to be made with respect to (1) the hierarchical structure of the estimation and, related to this, (2) the use of random effects for each observation. The nature of "hierarchical" models is that the DGP is assumed to be split into multiple levels: the observations are generated by some function of parameters which are themselves generated by some function of "hyper"-parameters. A straightforward justification of this structure is when variables that inform our estimation problem are recorded at different levels (like voters with certain characteristics, within districts with certain characteristics, within states with certain characteristics). It is also specifically helpful for count problems when overdispersion is present—that is, the variance of a count process is greater than its mean (Gill 2002, 351). Due to the multiple levels of data generation, hierarchical models typically represent each observation as the product of a related but different DGP. As a result, the observations at the lowest levels—that is, the actual observations of the LDV of interest—are treated as *de facto* random effects.[3]

---

[2]We give some sample code in the Appendix that users may find helpful for replication.

[3]Hierarchical treatments also allow researchers to move away from problematic small sample properties of some LDV estimators: for example, logistic regression standard errors are misleading in small samples (e.g., Davison 2003, 488–90).
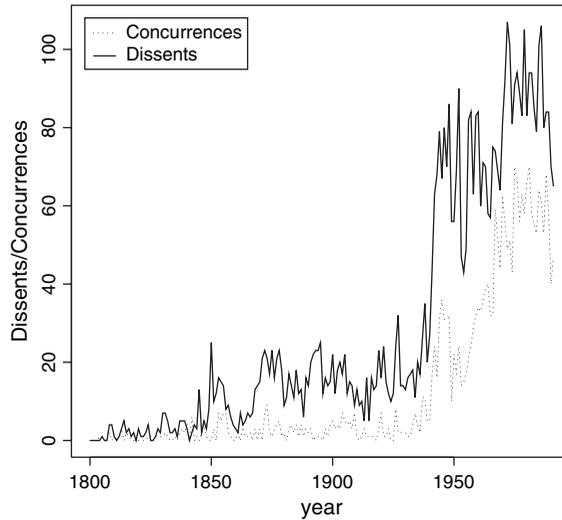
Fig. 1 Dissents and concurrences on Supreme Court, by year.

Using a hierarchical specification, though natural in some instances, is not required for Bayesian LDV change point models.[4] In our first example below, we use such a specification, but in the second and third, we do not.

3.2 *Rough Justice(s)*

The Justices of the Supreme Court of the United States generally accord with one another in reaching their decisions but not as often as they used to. If, indeed, a Justice wishes to make his dissent from the majority opinion public, he can do one of two things. First, he may file a "dissenting opinion," explicitly disagreeing with the *outcome* and rationale of the majority belief. Or, he could issue a "concurrence" with the majority opinion that states or clarifies his differing *rationale*. For the 19th century, several scholars have argued that a "norm of consensus" pervaded proceedings, which reflected a belief that unanimity strengthened public perceptions of the Court. By contrast, the second half of the 20th century is seen as a period of relative "dissensus" (Epstein, Segal, and Spaeth 2001, 362–3). Figure 1 makes this point graphically.

Simply stated, somewhere between 1940 and 1950, there was an explosion of dissent which, by the close of the century, had not returned to its prewar levels. The theories put forward to explain this pattern have been varied, though several concentrate on the ascendancy of Harlan Fiske Stone to the position of Chief Justice in 1941 (Epstein, Segal, and Spaeth 2001). By contrast, Caldeira and Zorn (1998) argue that there was basically no single break in norms at all and that both series characterized by shifting consensus levels in their entirety.

3.2.1 Poisson structural break modeling

The underlying model for this example is both well known and widely used (see, Gill (2002, 357) and the sources cited there for examples). Dissent and concurrence numbers

---

[4]Nor, indeed, do hierarchical models need to be Bayesian, though they most often are; for example, a prior distribution is specified for the highest level of the DGPs in a way that would be unacceptable to frequentist scholars.

are count data, so it makes sense to model the outcome Poisson with gamma random effects. As alluded to above, this will take care of overdispersion and makes the model a *de facto* negative binomial. A convenient choice for the prior on the arrival rate of the Poisson is the gamma which is conjugate with the Poisson. As noted above, in packages like *BUGS* such conjugacy is not strictly required, though it can speed computation. The gamma is characterized by α and β, a shape and scale parameter, which are typically given gamma hyperparameters or hyperpriors (hyperparameters in this case).

The statistical model is represented as follows (for the two series separately). If there is a change point, then we are asserting the existence of two Poisson DGPs. In particular,

$$
\begin{aligned}
Y_i |\ \theta \sim \mathcal{P}(\theta), \quad i = 1, \ldots, k, \\
Y_i |\ \eta \sim \mathcal{P}(\eta), \quad i = k+1, \ldots, T,
\end{aligned}
\tag{4}
$$

where Y_i indexes the series of count data, θ gives the Poisson arrival (λ) rate prior to k and η gives the arrival rate after k. The statistical model begins with equation (4) and is completed with

$$
\begin{aligned}
\theta_i &\sim \mathcal{G}(\alpha, \beta), \\
\eta_i &\sim \mathcal{G}(\gamma, \delta), \\
\alpha &\sim \mathcal{G}(A, B), \\
\beta &\sim \mathcal{G}(C, D), \\
\gamma &\sim \mathcal{G}(E, F), \\
\delta &\sim \mathcal{G}(G, H), \\
k &\sim (\text{discrete})\mathrm{U}[1, 2, \ldots, T],
\end{aligned}
\tag{5}
$$

where the Y_i are assumed conditionally independent and α and β are assumed independent. The letters A through H here represent the parameters of the gamma distributions "higher up" the hierarchy of the model. Notice that the arrival rates are properly indexed by i: this characteristic parameter is now able to take different values for every year of the data.

3.2.2 Results

A 100,000 iteration burn-in was judged sufficient for convergence, and a further 100,000 iteration yielded the results in Table 1.[5]

In general, the results here accord with our intuitions since the change point is 1941. Recall that this break has been properly modeled as a count process; it is possible to graphically display the (estimated) Poisson densities on either side of the change point, using the arrival rate from Table 1. Figure 2 does just that for the dissents series.

3.3 *(Ex)change Points*

Roughly speaking, for developed countries, there have been two exchange rate regimes in the postwar period (Bordo and Schwartz 1999): the first, from 1946–1971, was the Bretton Woods system by which each signatory country adopted a monetary policy that maintained the exchange rate of its currency within a fixed value—±1%—in terms of gold. The

---

[5]In general, the posteriors were (relatively) smooth and unimodal. Convergence assessed using Geweke (1992) and Raftery and Lewis (1995) diagnostics. Absolute value of G is less than two. Also, Raftery and Lewis (1992) diagnostic satisfied at conventional levels.

Table 1 Change point model results for concurrences and dissents from Supreme Court majority opinion [95% credible interval]

	Dissents		Concurrences	
	First period	Second period	First period	Second period
k	—	1941 [1941, 1941]	—	1941 [1941, 1941]
$\bar{\theta}$ (mean)	10.3 [7.2, 14.0]	—	2.0 [1.2, 3.0]	—
$\bar{\phi}$ (mean)	—	74.4 [64.0, 86.7]	—	38.8 [33.7, 44.2]

second period of postwar exchange regimes began in the March of 1973 with generalized floating rates. Initially, "dirty" floats were the norm, whereby monetary authorities intervened regularly to alter both the level of volatility and the exchange rates themselves. By the 1990s, it had evolved into a system of freer floats with intervention increasingly rare. In a series of works, Bordo and Flandreau (2001, 2003) argue that this logic applies somewhat differently for developing countries in the "periphery" of the world economy. Initially not invited into the Bretton Woods arrangement, they have subsequently opted for fixed rates to avoid (foreign) capital flight in the event of a depreciation.

Commensurate with this logic, in our illustration here, we study a panel of 21 countries making a binary exchange rate regime choice (fixed or floating) with observations recorded between the first quarter of 1959 and the last quarter of 1996.[6] One way to operationalize "development" is to consider Gross National Product (GNP) as a proxy and utilize it as a regressor in the application.

3.3.1 Logit structural break modeling

If there is a structural break, then we are asserting the existence of two Bernoulli DGPs. In particular,

$$
\begin{aligned}
Y_i|\ \theta &\sim \mathcal{B}(\theta), \quad i=1,\ldots,k, \\
Y_i|\ \eta &\sim \mathcal{B}(\eta), \quad i=k+1,\ldots,T,
\end{aligned}
\tag{6}
$$

where Y_i indexes the series of binary data, T is the number of observations, θ gives the binomial "success" rate prior to k, and η gives the success rate after k. We utilize a logit for the binary Y_i and the statistical model is completed with

$$
\begin{aligned}
\theta_i &= [1 + \exp(-\mathbf{X}_i^{\theta}\beta_i^{\theta})]^{-1}, \\
\eta_i &= [1 + \exp(-\mathbf{X}_i^{\eta}\beta_i^{\eta})]^{-1}, \\
\beta_i^{\theta,\eta} &\sim \mathcal{N}(0,A), \\
k &\sim (\text{discrete})\ \mathrm{U}[1,2,\ldots,T],
\end{aligned}
\tag{7}
$$

where A is some large number.

Notice that the model is estimated in the following fashion: the countries are treated as a cross-section by time period from which the coefficient estimates are then calculated.

---

[6]Countries are United States, United Kingdom, Austria, Belgium, Denmark, France, Germany, Italy, Netherlands, Norway, Sweden, Switzerland, Canada, Japan, Finland, Greece, Ireland, Portugal, Spain, Australia, and New Zealand. Data available from Andrew Rose: http://faculty.haas.berkeley.edu/arose/RecRes.htm
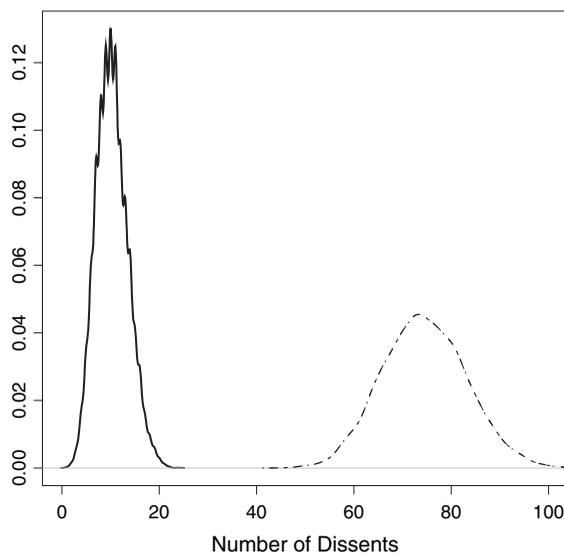
Fig. 2 Estimated densities for Supreme Court dissent counts before (solid line) and after (broken line) 1941 change point.

Hence, there are T (where T represents the number of quarters, not the total number of observations) estimates of the parameters from which a value for k is computed. This has the consequence that a probability of success is calculated for every $i \in T$, rather than for every $i \in T$, $j \in M$, where M is the number of countries in the panel. This is admittedly a slightly cruder treatment than might be expected, but it makes the computational burden somewhat lighter.

For the moment, denote posteriors with the notation Pr() and notice that the joint posterior for this problem is

$$\Pr(\theta, \eta, k \mid \mathbf{Y}) \propto \mathcal{L}(\theta, \eta, k \mid \mathbf{Y}) p(\theta \mid \beta_i^\theta) p(\eta \mid \beta_i^\eta) \tau(k)$$

$$= \left[\prod_{i=1}^{k} \theta^{Y_i} (1-\theta)^{1-Y_i} \right] \times \left[\prod_{i=k+1}^{T} \eta^{Y_i} (1-\eta)^{1-Y_i} \right]$$

$$\times \left[(2\pi\sigma_A^2)^{-\frac{1}{2}} \exp\left(\frac{-(\theta - \mu_{\beta_\theta})^2}{2\sigma_A^2} \right) \right]$$

$$\times \left[(2\pi\sigma_A^2)^{-\frac{1}{2}} \exp\left(\frac{-(\eta - \mu_{\beta_\eta})^2}{2\sigma_A^2} \right) \right] \times \frac{1}{T}, \tag{8}$$

where β is subscripted to refer to the mean of the relevant parameter and σ_A^2 refers to the variance A. Here, π actually refers to the constant $3.142\cdots$.

At least two comments are in order here. First, despite the fact that there are presumably country-specific effects, they are not being accounted for explicitly. The "second" level is left out here for reasons of parsimony: it makes the model easier both to describe and conceptualize and to estimate. Note that, in theory at least, it is trivial to reestimate the model with country-specific effects.

Table 2 Coefficients [95% credible interval] for logit change point model predicting fixed-rate choice, 1959–1996

	Standard logit model	Change point model	
		First regime	Second regime
Change point (k)	—	—	1971.1 [1971.1, 1971.1]
Intercept	4.43 [3.98, 4.86]	−0.719 [−1.987, 0.394]	2.293 [−1.047, 3.606]
GNP	−4.673 [−5.158, −4.127]	10.94 [8.426, 14.02]	−12.59 [−15.62, −10.01]

The second point pertains to the relation between the observations themselves (both temporally and spatially). An obvious concern that arises with nearly all panel (time series, cross-section) work is the independence (and identical distribution) of the observations and the methods for dealing with data that do not meet this condition. As currently written, this model is not hierarchical in the conventional sense. Nonetheless, if researchers were willing to pay the extra computational burden, the hierarchical model (with random effects) would not require i.i.d., but only "exchangeability." The basic idea here is that the observations we have are generated in the same way for every data value—conditional, obviously, on the particular parameter values. Crudely, it may be interpreted intuitively as meaning that countries are genuinely comparable units, insofar as the exchange rate regime choice is produced by the same causal mechanism for every country.[7]

3.3.2 Results

Convergence was judged to have occurred after 100,000 iterations,[8] and a further 500,000 post–burn-in iterations yielded the results in Table 2.

These results look promising. Notice first that the algorithm finds very strong evidence for a change point in period 49. This corresponds to a break after the first quarter of 1971, which therefore includes the August move by President Nixon to take the United States off gold, and thereby end the Bretton Woods agreement.[9] The signs on the coefficients are also encouraging, given the thesis above: before Bretton Woods ended, wealthier nations (with higher GNPs) fixed their currencies. By contrast, the aftermath saw developed (by proxy, more financially mature) nations opt for flexible rates, although a lower GNP was associated with a fixed regime choice.

Though these results are not a direct verification of the Bordo and Flandreau (2003) explanation of regime choice, they are certainly in line with the central argument presented there. Finally, it is worth noting that the strategy of simply running a standard logit model without structural break (see the first column in the table) yields a negative coefficient. Hence, a change point ignorant approach will imply that, from 1959 onward, wealthy countries (with higher GNPs) tend to float their rates: yet this is, in fact, only true *after* the Bretton Woods period.

---

[7]See Gill (2002, 364–8) and Western (1998, 1242–3) for a more technically thorough discussion, with other examples.

[8]All posteriors were (relatively) smooth and unimodal. Convergence of the change point parameter assessed using Geweke (1992) and Raftery and Lewis (1995) diagnostics. Convergence for the coefficient estimates was less easily obtained, but varying the priors somewhat resulted in very similar posterior findings.

[9]The change point posterior density is degenerate for period 49; hence, the posterior odds ratio for no change is zero.

Table 3 Fourth versus Fifth Republic government durations

	Number of governments	*Mean duration*
Fourth Republic	22	199
Fifth Republic	18	951

3.4 *Vive La Difference!*

For Huber (1996, 1),

> [t]he transition from the Fourth to the Fifth Republic in France provides what may be the most
> dramatic historical example of how changing the rules of a democracy can change the performance
> of a democracy.

The Fourth Republic itself was ripe for reform; it had exhibited both executive impotence and executive instability. In this section, the concentration is on the latter. Interested readers can consult Huber's text on the precise nature of the changes to the constitution in 1958, but essentially there are "two types of rules" that characterize the Fifth Republic (Huber 1996, 2). First, the position of President of France was made substantially stronger. Now able to stand above and apart from the political scrum of the legislature, France's head-of-state could more easily hire and fire prime ministers. Second, the post-reform government almost always enjoys a majority of support in the legislature, a movement toward a situation generally referred to as *le parlementarisme rationalisé*.

To summarize the magnitude of the change that these reforms wrought, Table 3 simply presents the number of separate prime ministerial tenures from 1947 through to November 2005 demarcated by Republic and gives the mean duration times in days. Clearly, governments last longer than they used to.

In fact, the first column of Table 3 is deceptive. Rather than looking at the mean duration of prime ministers, a better measure of stability is to study the regime life of different *control* of the premiership. France is not, and has never been, a "Westminster" system: the prime minister is far from the all-powerful head-of-government with guaranteed election-to-election tenure, that one might characterize (or perhaps caricature) him as in a state like Britain or Canada. Since the prime minister is—in modern times—beholden to the President's whim, it makes more sense to concentrate on party control of the head of the cabinet. *Ipso facto*, Fig. 3 gives a graph of *party* tenure of the Premiership in France: the leap in tenure around 1958—when the broken line cuts the curve—should be self-evident.

To clarify the coding here, from May 10, 1988, until March 31, 1993, the Prime Minister was a socialist, though three different politicians (Michel Rocard, Edith Cresson Pierre Bérégovoy) held the position. For current purposes, this "counts" as one government of 1784 days. By contrast, the period February 17, 1955, through June 12, 1957, is three separate governments, since the party control of the office changed from the *Section Française de l'Internationale Ouvrière* (SFIO) to the Radicals to the SFIO once again. Notice then that a long *Union pour la nouvelle République* (UNR) hold on the Premiership, beginning with De Gaulle in 1958, straddles the creation of the Fifth Republic.

3.5 *Exponential Structural Break Modeling*

This section operationalizes "stability" with duration modeling (Cioffi-Revilla 1984; Lijphart 1984). The "duration" in question is the survival period of the party in government.
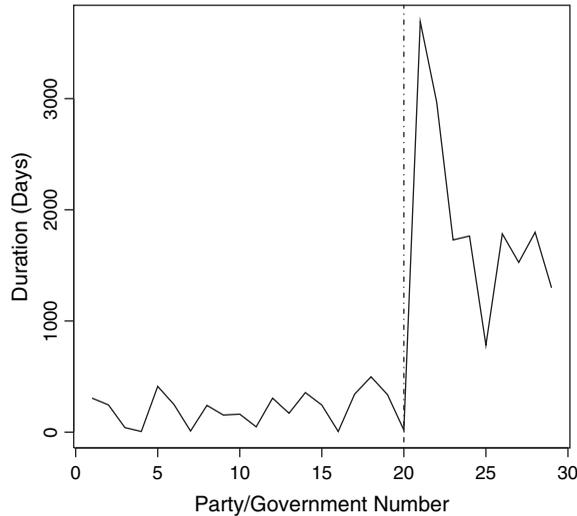
Fig. 3 French Party Tenure of Premiership, 1947–2005.

Modeling durations is certainly not new to this paper: King et al. (1990) analyze the survival of 314 European cabinet governments, using a "unified" exponential model. It is "unified" in the sense that "attributes" theory and "events process theory" are combined in one statistical approach. The former approach, exemplified by Strom (1985), is that cabinet duration is accounted for by "attributes" or "properties": these may derive from the (contemporaneous) political system, the party system, or the particular cabinet in question (King et al. 1990, 847). By contrast, the other approach perceives the fall of each government as being "generated by a particular 'critical' or 'terminal' event. By making assumptions about the probability of such events, the events theorists model the pattern of cabinet dissolution" (King et al. 1990, 847). The model used below is exponential, though there are no covariates entering the hazard rate: hence it is something of a hybrid.[10]

3.5.1 Model

The statistical model is represented as follows. If there is a change point, then we are asserting the existence of two exponential DGPs. In particular,

$$
\begin{aligned}
Y_i| \, \theta &\sim \varepsilon(\theta), \quad i = 1, \ldots k, \\
Y_i \mid \eta &\sim \varepsilon(\eta), \quad i = k+1, \ldots, T,
\end{aligned} \tag{9}
$$

where Y_i indexes the series of survival times, θ gives the rate parameter of the exponential prior to k, and η gives the parameter after k. The model includes a constant, which is assumed normally distributed. Based on Huber's account, the theory is relatively sharp here, and we expect a "fixed" effect in each period with the durations in the early period shorter than those in the latter. Hence, we do not explicitly model every observation, and the random effects are dropped. The statistical model begins with equation (9) and is completed with

---

[10]If the researcher has a sufficiently sharp theory and the requisite data to hand, there is no theoretical bar to utilizing covariates. In the King et al. (1990) data itself, the record for France does not include the Fifth Republic.

$$\log(\theta_i) = \beta^\theta,$$
$$\log(\eta_i) = \beta^\eta,$$
$$\beta^{\theta,\eta} \sim \mathcal{N}(0, A), \tag{10}$$
$$k \sim (\text{discrete})\mathrm{U}[1, 2, \ldots, T],$$

where A is some large number.

For completeness, consider the joint posterior

$$
\begin{aligned}
\Pr(\theta, \eta, k \mid \mathbf{Y}) &\propto \mathcal{L}(\theta, \eta, k \mid \mathbf{Y}) p(\theta \mid \beta_i^\theta) p(\eta \mid \beta_i^\eta) \tau(k) \\
&= \left[\prod_{i=1}^{k} \theta \exp(-\theta Y_i) \right] \times \left[\prod_{i=k+1}^{T} \eta \exp(-\eta Y_i) \right] \\
&\quad \times \left[(2\pi\sigma_A^2)^{-\frac{1}{2}} \exp\left(\frac{-(\theta - \mu_{\beta_\theta})^2}{2\sigma_A^2} \right) \right] \\
&\quad \times \left[(2\pi\sigma_A^2)^{-\frac{1}{2}} \exp\left(\frac{-(\eta - \mu_{\beta_\eta})^2}{2\sigma_A^2} \right) \right] \times \frac{1}{T}, \tag{11}
\end{aligned}
$$

where β is subscripted to refer to the mean of the relevant parameter and σ_A^2 refers to the variance A. Here, π is once again the constant $3.142\cdots$.

3.6 *Results*

Three chains with different starting values were utilized. Convergence was judged to have occurred after 200,000 iterations,[11] and a further 200,000 postconvergence iterations produced the results in Table 4. The second and third columns give the estimates for the exponential parameter (λ) meaned out over the respective time periods ($1,\ldots,k; k+1,\ldots,T$) under the change point model and, for comparison, the first column gives the coefficients for a model with no change points.

The results here are quite encouraging.[12] Period (government) 20 is the long UNR administration that begins with De Gaulle as prime minister and includes the instigation (by De Gaulle himself) of the Fifth Republic. This fits with Huber's account and with intuitions about French politics. Notice that the no–change point model does not allow the researcher to explore the effects of the profound constitutional change that 1958 ushered in. The coefficients are the logged reciprocal hazard rates. Hence, in the first period, the mean duration of cabinets was $\frac{1}{\exp(0.54)} = 0.58$ years, or approximately 6–7 months. After the break, the duration was, on average, $\frac{1}{\exp(-2.22)} = 7.38$ years. For the no–change point model, the analyst would predict a mean duration of $\frac{1}{\exp(-0.726)} = 2.06$ or around 2 years.

To further explore the results, Fig. 4 uses the calculated rates of the exponentials from Table 4 to compare the cabinet durations graphically for the change point and no–change point models. The thick black line is the density function for the duration for the postbreak (Fifth Republic) governments; the tall dashed line is that for the prebreak (Fourth

---

[11]In general, the chains showed good mixing, and posteriors were (relatively) smooth and unimodal. Convergence assessed using Geweke (1992) and Raftery and Lewis (1995) diagnostics.
[12]The posterior odds ratio for no change is zero: there is no mass at $k = T$.

Table 4 Coefficients [95% credible interval] for exponential duration change point model, French prime ministerial data, 1947–2005

		Change point model	
	Standard duration model	First regime	Second regime
Change point (k)	—	—	De Gaulle (1958) [Faure (1955), De Gaulle (1958)]
log (λ)	−0.726 [−1.11, −0.377]	0.5416 [0.06517, 0.9728]	−2.216 [−3.055, −1.434]

Republic) governments; and the broken line between them is derived from the no–change point model. Notice the relatively smooth density with a long tail for the Fifth Republic.

4 Discussion

This paper is intended as a review—and to make political scientists aware—of some options for the LDV change point problem. We considered hierarchical models and applications for count, binary, and duration dependent variables and discussed various tradeoffs that are made when employing a Bayesian solution to the problem. Above we consider single change point problems, but future work in this area might fruitfully concentrate on multiple break points.

It would take relatively little manual effort to adjust the code used for the examples above to look for two or three or four change points. Unfortunately, there are computational issues, and for short data series—in the French Republic case, for example—estimating all the coefficients in question could become quite burdensome. But there is another philosophical issue. In the GLM context, we usually know how many coefficients we are looking for: if, say, we have three regressors and a constant predicting presidential turnout in a probit, we have an (estimated) $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$. The change point class of problems is slightly different in that we may be *uncertain* about how many parameters (breaks) actually characterize the time series process. In this paper, the question was crudely restricted to a binary option of "one" or "none." If we reprogram to look for four breaks, then we are committing to four as an upper bound and not assessing potential evidence for five, six, and so on.

Ultimately then, we would like a technique which enabled us to be *uncertain* over the number of breaks and allowed this to be *estimated*—albeit with prior information—along with our actual breaks and their effects (the coefficients). That is, we would like to know if the evidence suggests no breaks, one, two, three, or n. There seem at least two ways to proceed.

Chib (1998) suggests a Markov mixture model with each observation generated by latent state variables. This is a state space model, where both the model parameters and the probabilities that make up the transition matrix are given priors. For the latter priors, a Dirichlet process may be assumed, which allows the analyst some agnosticism regarding the parametric form required. In recent times, Park (2006) has applied exactly this method to the (American) presidential use of force in international relations.[13]

---

[13]See also Kim and Nelson (1999) and Nelson and Kim (1999) for examples of similar ideas applied to economics.
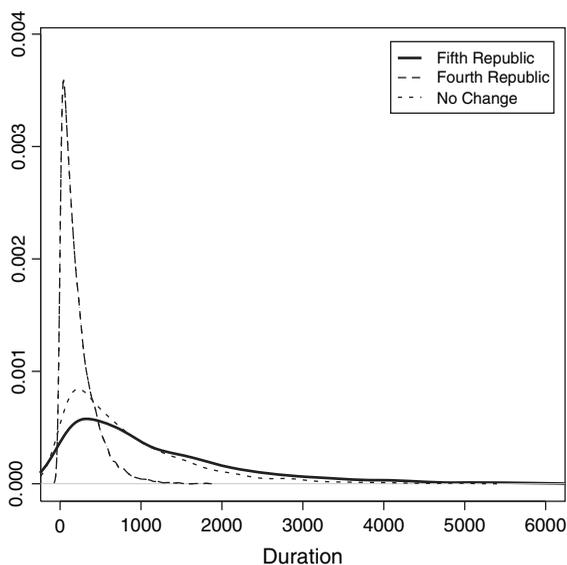
Fig. 4 Densities for estimated change point and no change point duration times.

Second, a "reversible jump" approach may be of interest. Green (1995) introduces a technique for simulating a posterior distribution whose dimensions are unknown. The general application is for cases where the number of parameters in the model is not *a priori* certain. In the change point case, this means that researchers could be uncertain about the number of breaks and yet still analyze the data at hand without commitment to strong priors. Green's (1995) method has not yet been seen in political science, which is unfortunate because other types of problems we regularly encounter—like cluster analysis—often involve some unknown number of parameters. We leave this possibility for future work.

Appendix A: Replication Information

Users can replicate the findings using any number of statistical packages with *R* and *winBUGS* being preferable. We follow *winBUGS* conventions for discrete priors on the change point given by "Stagnant: a changepoint problem" that can be found, for example, at http://mathstat.helsinki.fi/openbugs/data/Examples/Volumeii.html

A.1 *Poisson*

Beginning with the Poisson model of Section 3.2, the following winBUGS code is appropriate (data available at ICPSR, study 1142).

```
model{for(i in 1:N){
y[i]~dpois(lambda[i])

lambda[i]<-round(lam[i])

lam[i]<-eta[i]+ J[i]*zeta[i]
```

```
J[i] <- step(yr[i]-k-0.5) }

for(i in 1:N){
eta[i]~dgamma(alpha,beta)
zeta[i]~dgamma(gamma,delta)
}

alpha~dgamma(50,1)
beta~dgamma(50,1)
gamma~dgamma(50,1)
delta~dgamma(50,1)

k~ dcat(priort[])

for(i in 1:N){
            priort[i] <- punif[i]/sum(punif[])
            }
}

#data
list(y = c(
#data vector of concurrences here
),

punif=c( # vector of ones: one for each period in y
  ),

yr=c(
#vector of years: 1...192
),

N=192)

#inits
list(k=100)
```

A.2 *Logit*

Data may be obtained from http://faculty.haas.berekeley.edu/arose/RecRes.htm. Set up the data in the usual way for winBUGS—with x2 the vector of GNP values—and the following model code should produce the requisite results:

```
model{for(i in 1:N){
y[i] ~ dbern(p[i])
```

```
logit(p[i]) <- alpha0 + alpha[1]*x2[i] +
  beta0*J[i]+J[i]*beta[1]*x2[i]

J[i] <- step(yr[i]-k-0.5); }

for(i in 1:g){
            alpha[i] ~ dnorm(0.0, 1.0E-6)
            beta[i] ~ dnorm(0.0, 1.0E-6)
                }
for(i in 1:T) {
            priort[i] <- punif[i]/sum(punif[])
            }
alpha0 ~ dnorm(0.0, 1.0E-6)
beta0 ~ dnorm(0.0, 1.0E-6)
k~dcat(priort[]) }

#data list(y=c(#vector of y values here),
x2=c(
#vector of GNP values here
),
T=152,
g=1,
N=3192,

punif=c(
# vector of ones: one for each quarter
),

yr=c(
# vector of quarters (1 through 152) for each country:
)

#inits
list(alpha=c(0), beta=c(0), alpha0=0, beta0=0, k=20 )
```

A.3 *Duration*

The vector of durations is short (the figures refer to years), so we give the full data here:

```
list(y=c(0.838356164383562, 0.671232876712329,
  0.112328767123288,
0.0164383561643836, 1.12876712328767, 0.676712328767123,
```

```
0.0273972602739726, 0.66027397260274, 0.421917808219178,
0.443835616438356, 0.131506849315068, 0.838356164383562,
0.465753424657534, 0.975342465753425, 0.668493150684932,
0.0164383561643836, 0.936986301369863, 1.36438356164384,
0.926027397260274, 0.052054794520548, 10.1150684931507,
8.13424657534247, 4.73698630136986, 4.83287671232877,
2.14246575342466, 4.88767123287671, 4.18356164383562,
4.92602739726027, 3.55890410958904),
punif=c(1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,1),
T=29,
N=29,
yr=c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28,29))
```

The model is similar to the above:

```
model{
for(i in 1:N){
y[i] ~ dexp(m[i])

log(m[i])<-etaC+ zetaC*J[i]

J[i] <- step(yr[i]-k-0.5);
}

for(i in 1:N){
eta[i]~dnorm(0,0.001)
zeta[i]~dnorm(0,0.001)
}

etaC~dnorm(0, 0.001)
zetaC~dnorm(0,0.001)

for(i in 1:T) {
    priort[i] <- punif[i]/sum(punif[])

k~ dcat(priort[])
}

#inits
{
list(k=5)
```

```
list(k=20)
list(k=25)
}
```

Funding

Star Lab

References

Beck, Neal, Jonathan Katz, and Richard Tucker. 1998. Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science* 42:1260–88.

Bordo, Michael D., and Marc Flandreau. 2001. *Core, periphery, exchange rate regimes and globalization.* http://www.nber.org/papers/w8584. (accessed September 1, 2005).

———. 2003. Core, periphery, exchange rate regime and globalization. In *Globalization in historical perspective*, M. D. Bordo, A. M. Taylor, and J. G. Williamson, 417–72. Chicago, IL: University of Chicago Press.

Bordo, Michael D., and Anna J. Schwartz. 1999 Monetary policy regimes and economic performance: The historical record. In *Handbook of macroeconomics*, ed. J. Taylor and M. Woodford, 149–236. London: North-Holland.

Caldeira, 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.

Carlin, Bradley, Alan Gelfand, and Adrian Smith. 1992. Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics* 41:389–405.

Casella, George, and Edward I. George. 1992. Explaining the Gibbs sampler. *The American Statistician* 46:167–74.

Chib, Suddhartha. 1998. Estimation and comparison of multiple change-point models. *Journal of Econometrics* 86:221–41.

Cioffi-Revilla, Claudio. 1984. The political reliability of Italian governments. *American Political Science Review* 78:318–37.

Clarke, Kevin A. 2001. Testing nonnested models of international relations: Reevaluating realism. *American Journal of Political Science* 45:724–44.

Davison, A. C. 2003. *Statistical models.* Cambridge: Cambridge University Press.

Epstein, Lee, Jeffrey A. Segal, and Harold J. Spaeth. 2001. The norm of consensus on the US Supreme Court. *American Journal of Political Science* 45:362–77.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian statistics 4*, ed. J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger, 169–93. Oxford: Oxford University Press.

Gill, Jeff. 2002. *Bayesian methods: A social and behavioral sciences approach.* Boca Raton, FL: Chapman and Hall/CRC.

Green, Peter J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–32.

Greene, William H. 2003. *Econometric analysis.* New York: MacMillan Publishing Company.

Huber, John. 1996. *Rationalizing parliament: Legislative institutions and party politics in France.* Cambridge: Cambridge University Press.

Jackman, Simon. 2000. Estimation and inference via Bayesian simulation: An introduction to Markov chain Monte Carlo. *American Journal of Political Science* 44:375–404.

Kim, Chang-Jin, and Charles R. Nelson. 1999. Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics* 81:608–16.

King, Gary. 1988. Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model. *American Journal of Political Science* 32:838–63.

King, Gary, James E. Alt, Nancy Elizabeth Burns, and Michael Laver. 1990. A unified model of cabinet dissolution in parliamentary democracies. *American Journal of Political Science* 34:846–71.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science* 44:341–55.

Lijphart, Arend. 1984. A note on the meaning of cabinet durability. *Comparative Political Studies* 17:163–66.

Mueller, John. 1991. Changing attitudes towards war: The impact of the first world war. *British Journal of Political Science* 21:1–28.

Nelson, Charles, and Chang-Jin Kim. 1999. *State-space models with regime switching: Classical and Gibbs sampling approaches with applications.* Cambridge, MA: MIT Press.

Oneal, John R., and Bruce Russett. 1997. The classical liberals were right: Democracy, interdependence, and conflict, 1950–1985. *International Studies Quarterly* 41:267–94.

Park, Jong Hee. 2006. Bayesian analysis of structural changes: Historical changes in US Presidential Uses of Force Abroad. Paper presented at the American Political Science Association Meeting, August 2006, Philadelphia, PA.

Raftery, A. E., and S. M. Lewis. 1992. How many iterations in the Gibbs sampler? In *Bayesian Statistics 4*, ed. J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger, 763–73. Oxford: Oxford University Press.

Ruckman, P. S. 1997. Executive clemency in the United States: Origins, development, and analysis (1900–1993). *Presidential Studies Quarterly* 27:251–71.

Strom, Kaare. 1985. Party goals and government performance in parliamentary democracies. *American Political Science Review* 79:738–54.

Tate, Katherine. 1991. Black political participation in the 1984 and 1988 presidential election. *American Political Science Review* 85:1159–76.

Western, Bruce. 1998. Causal heterogeneity in comparative research: A Bayesian hierarchical modelling approach. *American Journal of Political Science* 42:1233–59.

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

Wood, B. Dan. 2000. Weak theories and parameter instability: Using flexible least squares to take time varying relationships seriously. *American Journal of Political Science* 44:603–18.