Political Analysis (2004) 12:354–374
doi:10.1093/pan/mph023

A Bayesian Change Point Model for Historical Time Series Analysis

Bruce Western and Meredith Kleykamp
Department of Sociology, Princeton University, Princeton, NJ 08544
e-mail: western@opr.princeton.edu

Political relationships often vary over time, but standard models ignore temporal variation in regression relationships. We describe a Bayesian model that treats the change point in a time series as a parameter to be estimated. In this model, inference for the regression coefficients reflects prior uncertainty about the location of the change point. Inferences about regression coefficients, unconditional on the change-point location, can be obtained by simulation methods. The model is illustrated in an analysis of real wage growth in 18 OECD countries from 1965–1992.

1 Introduction

Conventional statistical models fail to capture historical variability in political relationships. These models are poorly suited for analyzing periods in which there are basic shifts in institutions, ideas, preferences, or other social conditions (Büthe 2002; Lieberman 2002). Despite the limits of conventional models, the importance of historical variability in political relationships is widely claimed. Historical institutionalists in comparative politics offer some substantive motivation: "Political evolution is a path or branching process and the study of points of departure from established patterns becomes essential to a broader understanding of political history" (Thelen and Steinmo 1992, p. 27). Lieberman (2002, p. 96) makes a similar point: "What we are after is an explanation not of ordinary predictable variation in outcomes but of extraordinary change, where relationships among explanatory factors themselves change."

Statistical analysis can begin to meet this challenge by examining temporal instability in quantitative relationships. In the analysis of regression models, temporal instability is studied in two main ways. First, researchers offer explicit theories about the timing of structural breaks in regressions (e.g., Schickler and Green 1997; Mitchell et al. 1999; Richards and Kritzer 2002). Second, an influential sociological paper by Isaac and Griffin (1989) rejects "a-historicism in time series analysis" in favor of an historically sensitive approach in which structural instability is an ever-present possibility.

Specific theories of the timing of structural breaks lead to formal tests of change points, while the "historical time series analysis" of Isaac and Griffin (1989) takes a diagnostic

Authors' note: This paper was prepared for the annual meetings of the Social Science History Association, St. Louis, Missouri, October 2002. Research for this paper was supported by a grant from the Russell Sage Foundation and by National Science Foundation grant SES-0004336. We thank *Political Analysis* reviewers, Bob Erikson, and Jeff Gill for helpful comments on earlier drafts. Computer code for the analysis in this paper is available on the *Political Analysis* Web site.

approach. Where specific theories are proposed, a change point in the regression is usually specified a priori and modeled with a dummy variable for time points after the change point. The change-point dummy might be used to describe a mean shift in the dependent variable or to allow for a change in the effects of covariates. For example, Richards and Kritzer (2002) argue that U.S. Supreme Court decisions are shaped by historically variable jurisprudential regimes. A regime of content neutrality in freedom of expression cases was established by several decisions in 1972. The effects of attitudes of the justices on free expression cases are smaller after the establishment of the new regime. Inference about the difference in effects before and after 1972 is provided by a chi-square test. In contrast, the diagnostic approach proceeds without a formal test of a parametric model for a change point. Structural instability is detected, for example, with plots of regression coefficients estimated for data within a window of time that moves along the series (Isaac and Griffin 1989, p. 879).

Formal parametric tests and diagnostic methods each have different limitations. Say we test for a mean shift in some outcome at time θ. Data indicating a break before θ should count against the hypothesis, but the observed difference in means before and after θ may lead us to find in favor of our hypothesis. Formal tests at least provide an inference about a change point. Searching for change points with diagnostics risks mistaking random variation for structural change. Also, if a model is fit after diagnostics are used to locate a change point, conventional t statistics and p values do not reflect prior uncertainty about the timing of structural shifts.

This paper describes a simple Bayesian model for change points that combines the advantages of diagnostic and parametric approaches but addresses their limitations. Parametric models assume the location of a change point, but diagnostic methods allow the timing of change to be discovered from the data. Like diagnostic methods, the Bayesian analysis treats the timing of change as uncertain and the location of a change point as a parameter to be estimated. This approach allows evidence for a change before a hypothesized date to count against the hypothesis. Like parametric models, the Bayesian model yields statistical inferences about regression coefficients. However, these inferences reflect prior uncertainty about the location of the change point that is unaccounted for in conventional models.

The Bayesian change-point model can be estimated using the Gibbs sampler. Although the model is relatively simple, estimation is computer intensive and requires some programming facility. We thus describe a simple alternative method that can, in part, be calculated using standard regression output.

2 An Ad Hoc Method for Studying Change Points

If we think that the effects of causal variables change over time, we can model this with a dummy variable that takes the value of zero up to the time point marking the end of the first regime, and the value of one thereafter. For example, Isaac and Griffin (1989) studied the relationship between strikes and labor union density in the United States. They argued that the relationship between strikes and unionization changed after the Supreme Court upheld the National Labor Relations Act in 1937. To capture this changing relationship, we could define a dummy variable I_t that equals 0 for each year up to 1936 and 1 for years 1937 and following. A model for the dependence of strikes, S_t, on unionization, U_t, would then be written:

$$S_t = \beta_0 + \beta_1 U_t + \beta_2 I_t + \beta_3 I_t U_t + e_t,$$

where e_t is an error term. Before 1937, the effect of unionization on strikes is given by the regression coefficient, β_1. After 1937, the effect is $\beta_1 + \beta_3$.

We can write the dummy variable more generally as a function of the change-point year, $I_t(1937)$. Even more generally, we can replace a specific change point with a variable, θ, whose value is not yet specified. In this case, the Isaac-Griffin model would be written:

$$S_t = \beta_0 + \beta_1 U_t + \beta_2 I_t(\theta) + \beta_3 I_t(\theta) U_t + e_t.$$

In addition to the regression coefficients, the model now has the change point, θ, to estimate.

How can we estimate the location of the change point? A simple method tries a range of values for θ and examines the model's goodness of fit. We could begin by assuming that the change point for the Isaac-Griffin model came in 1915 rather than 1937. Under this specification, the variable $I_t(\theta = 1915) = 0$ for years before 1915 and $I_t(\theta = 1915) = 1$ for years 1915 and after. We fit a model with this period effect and record the R^2 statistic. Next we try $\theta = 1916$ and so on. Calculating R^2 for $\theta = 1915, 1916, \ldots, 1950$ yields a sequence of R^2 statistics that can be plotted. We can then choose the value of θ with the maximum R^2.

We illustrate this approach with some artificial data where the location of a change point is known. Two time series of length 100, x_t and y_t, are generated. A set of fixed values is chosen for x_t. The dependent variable, y_t, is generated according to

$$y_t = .5 + .5x_t + .5I_t(37) + .4I_t(37)x_t + e_t,$$

where $I_t(37)$ is a dummy variable that equals 0 up to $t = 36$, and 1 from $t = 37, \ldots, 100$, and e_t consists of 100 random deviates generated from a normal distribution with mean 0 and standard deviation .5. With these artificial data, the change point $\theta = 37$. The data are shown in Fig. 1. The time series (top panel) provide a small suggestion of a change point at $t = 37$, but a clearer picture is given by the scatterplot in the lower panel. The scatterplot shows that the regression line for the first regime is supported mostly by low values of x_t, while the regression line for the second regime is supported mostly by large values of x_t. The difference in slopes is relatively small, but the difference in intercepts is clearly indicated.

Implementing our ad hoc method to estimate the change point, we fit 81 regression models with ordinary least squares (OLS), trying change points from $\theta = 10, \ldots, 90$. These 81 models yield a time series of R^2 statistics (Fig. 2). The maximum R^2 comes at $\theta = 41$, close to, but not exactly equal to, our known change point of $\theta = 37$.

This ad hoc method of searching for the best-fitting change point has a good statistical justification. Assuming that the error term is normally distributed, the time series of R^2 statistics is proportional to the profile log likelihood for θ. The time point of maximum R^2 ($t = 41$ for our artificial data) is the maximum likelihood estimate of θ. Defining the dummy variable and fitting the time series is easily done using standard statistical software. Thus maximum likelihood estimation of θ is simple, if somewhat laborious.

Having estimated the change point, θ, how do we estimate the regression coefficients? A naive approach estimates the coefficients conditional on the maximum likelihood estimate of θ. After estimating $\hat{\theta} = 41$, say, we can calculate the regression coefficients, their standard errors, and p values using OLS. Conventional standard errors and p values

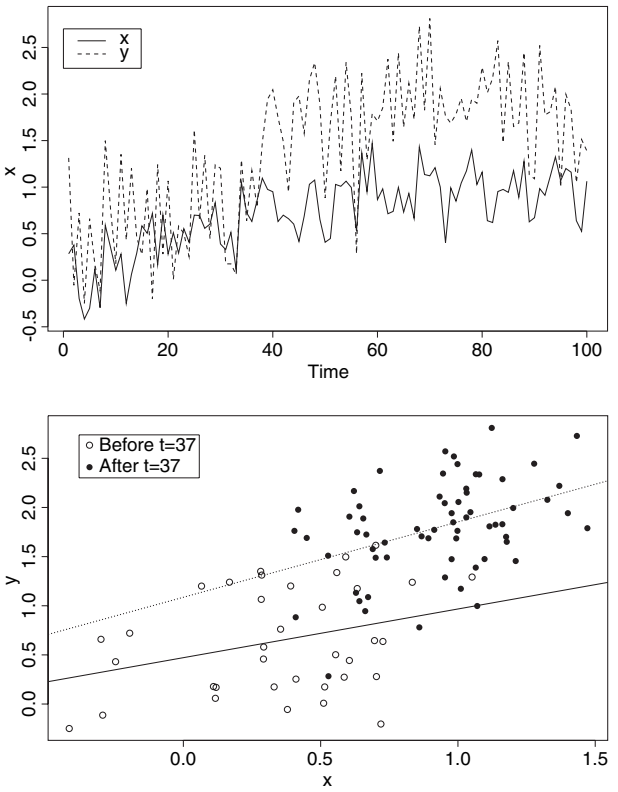

Fig. 1 The top panel shows 100-point time series for the artificial data, x and y, where a change in the relationship between the variable occurs at time point $t = 37$. The bottom panel is a scatterplot of x against y, with two regression lines fitting the time points $t < 37$ (open circles) and $t \geq 37$ (filled circles).

will be too small, however, increasing the risk of an incorrect inference of statistically significant effects. These inferences are optimistic because the model assumes with certainty that $\theta = 41$ for the purposes of calculating the coefficients, but θ itself is uncertain and estimated from the data. By using the data to search for the best-fitting model and to estimate that model's coefficients, significant results are more likely due to random variation than conventional p values suggest (Freedman 1983). From a Bayesian perspective, inferences about the coefficients are optimistic because we have not correctly accounted for our uncertainty about θ. In sum, the ad hoc method can be used for maximum likelihood estimation of θ but should should not be used to estimate regression coefficients because it ignores uncertainty about the location of the change point. The Bayesian model averages over this uncertainty to obtain correct inferences about the coefficients.

3 The Bayesian Change Point Model

Consider a regression model for the dependent variable, y_t, of the form

$$\hat{y}_t = \beta_0 + \beta_1 x_t + \beta_2 I_t(\theta)_t + \beta_3 I_t(\theta)_t x_t \quad t = 1, \ldots, T,$$

or equivalently in matrix notation,

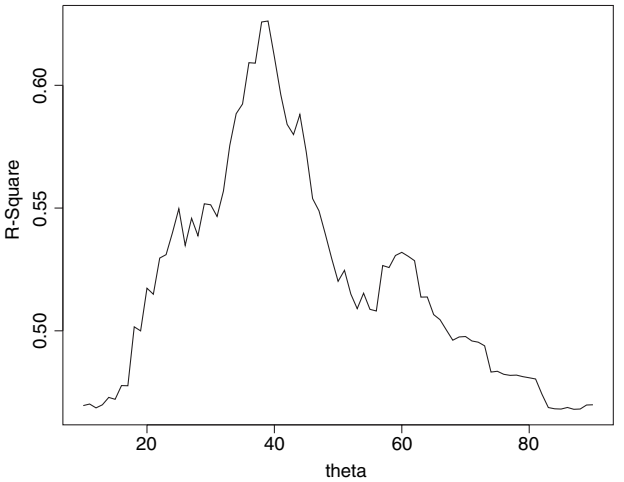

Fig. 2 Time series of R^2 statistics, from regressions on artificial data for x and y, $\theta = 10, 11, \ldots, 90$.

$$\hat{\mathbf{y}} = \mathbf{X}_\theta \boldsymbol{\beta}, \tag{1}$$

where the change point indicator $I_t(\theta) = 0$ for $t < \theta$ and $I_t(\theta) = 1$ for $t \geq \theta$, $\mathbf{y}$ is a vector of observations on the dependent variable, the matrix $\mathbf{X}_\theta$ includes all regressors, and $\boldsymbol{\beta}$ is a vector of regression coefficients. We write $\mathbf{X}_\theta$ as a function of θ because different change points will yield different regressors. Assuming that $\mathbf{y}$ conditionally follows a normal distribution, the contribution of an observation, y_t, to the likelihood of θ is

$$L_t(\theta; y_t) \propto \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_t - \hat{y}_t)^2}{2\sigma^2}\right], \tag{2}$$

where the likelihood depends on θ through the regression function in Eq. (1). For a given change point, say $\theta = k$, the likelihood evaluates to $L(\theta = k; \mathbf{y}) \propto \prod_t L_t(\theta = k; y_t)$. Here the error variance is written as σ^2, but we shall work with the precision $\tau = \sigma^{-2}$. Because the likelihood function for this model is discontinuous for discrete values of θ, it is difficult to obtain unconditional inferences about the regression coefficients, $\boldsymbol{\beta}$, using standard likelihood methods.

3.1 *Prior Distributions for the Change Point Model*

We can obtain inferences about the change point and the coefficients using a Bayesian approach that specifies prior distributions for the parameters. The priors and the likelihood can be written as follows:

$$\begin{aligned} p(\tau) &= \text{Gamma}(n_0, s_0) \\ p(\boldsymbol{\beta}) &= N(\boldsymbol{\beta}_0, \mathbf{V}_0) \\ p(\theta) &= (T-1)^{-1}, \quad \theta = 1, \ldots, T-1 \\ p(\mathbf{y} \mid \mathbf{X}_\theta) &= N(\hat{\mathbf{y}}, \tau^{-1}), \end{aligned}$$

where Gamma(a, b) is a gamma distribution with shape parameter a and expectation a/b. The prior for θ is a discrete uniform distribution that allocates equal prior probability to

each time point. A noninformative prior for τ sets n_0 and s_0 to small positive numbers, say .001. A noninformative prior for the coefficients sets $\boldsymbol{\beta}_0 = \mathbf{0}$ and $\mathbf{V}_0$ to a diagonal matrix with large prior variances, say 100.

Estimation of this model can be approached in three ways. Full conditional posterior distributions can be used to form a Gibbs sampler for posterior simulation. Alternatively, Chin Choy and Broemeling (1980) derive an expression for the marginal posterior of θ. Even more simply, inference for θ and the coefficients might be based on the profile log likelihood, a function of the residual sums of squares from OLS regressions.

3.2 *The Gibbs Sampler*

The Gibbs sampler is a method for Bayesian estimation that simulates draws from the posterior distribution. To implement the Gibbs sampler, posterior distributions are specified for each parameter conditional on all the other parameters in the model. Sampling from these full conditional posterior distributions ultimately yields draws from the unconditional posterior distribution (e.g., Jackman 2000). In principle, the Gibbs sampler is an extremely flexible tool for Bayesian inference allowing the estimation of any model for which full conditional posteriors can be specified.

The Gibbs sampler begins by setting initial values of the change point, the precision, and the regression coefficients. The sampler randomly draws one of these parameters—written θ^*, τ^*, and $\boldsymbol{\beta}^*$—conditional on current values for the others:

1. Draw the precision τ^* from Gamma$(n_0 + n/2, s_0 + SS^*/2)$, where the current value of the sums of squares $SS^* = \sum e_t^*$, and $e_t^* = y_t - \mathbf{x}'_{\theta^* t}\beta^*$.
2. Draw the vector of regression coefficients from the multivariate normal distribution $N(\boldsymbol{\beta}_1, \mathbf{V}_1)$, with mean vector $\boldsymbol{\beta}_1 = \mathbf{V}_1(\mathbf{V}_0\boldsymbol{\beta}_0 + \tau^*\mathbf{X}'_{\theta^*}\mathbf{y})$, and covariance matrix $\mathbf{V}_1 = (\mathbf{V}_\mathbf{0}^{-\mathbf{1}} + \tau^*\mathbf{X}'_{\theta^*}\mathbf{X}_{\theta^*})^{-1}$.
3. Draw the change point from the discrete distribution $p(\theta = t \,|\, \mathbf{y}) = L^*(\theta = t; \mathbf{y}) / \sum_t L^*(\theta = t; \mathbf{y})$, where the likelihood in Eq. (2) is evaluated at each time point $\theta = 1, \ldots, T - 1$ using current values of the parameters $\boldsymbol{\beta}^*$ and τ^*.

The algorithm is termed a blocked Gibbs sampler, because step 2 updates the block of all coefficients under all the change points. The Gibbs sampler can also accommodate a more general model in which a regression with no change point is included a priori. For this specification, $p(\theta \,|\, \mathbf{y})$ in step 3 also evaluates the likelihood of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$, including just the covariate x_t, setting coefficients $\beta_2 = \beta_3 = 0$.

In the simulation experiments below we obtained good performance for a small regression on one predictor with a burn-in of 100 iterations, and a sequence or chain of 2000 iterations. The method is computer intensive compared to some other Gibbs sampler applications because the probability of θ must be evaluated for every time point for every iteration. The Gibbs sampler is useful because it automatically provides inferences about the regression coefficients. The algorithm can also be generalized to introduce informative priors for the coefficients or θ if we have substantive knowledge about the effects or the location of the change point. Other likelihoods can also be specified for $\mathbf{y}$. For example, the Gibbs sampler could be used to detect structural breaks in event history models or in panel data with heterogeneous error variances. Results for an error-heterogeneity model with an informative prior on θ are reported in the application below.

A significant practical challenge for analysis with the Gibbs sampler involves assessing convergence of the algorithm. If the Gibbs chain is not stationary and the mean of the chain shifts as it moves over the parameter space, the algorithm may not adequately

sample from the unconditional posterior distribution. In practical applications, diagnostics should be used to assess convergence (Carlin and Louis 2002, pp. 172–183, review many of the options). Several parallel chains should also be run with widely spaced starting values. Because the posterior distribution for the change point has a simple form, alternative simulation methods are available that offer immediate convergence. These alternative methods stochastically sample from the posterior distributions of the coefficients, conditional on different change points.

3.3 *Stochastic Sampling from the Conditional Posterior*

The Gibbs sampler is computationally intensive, requiring thousands of iterations, but the posterior distribution of θ can be calculated exactly and quickly. Although analytical results do not provide an expression for the unconditional posterior distribution of the regression coefficients, the conditional posteriors given θ are known to be multivariate t whose means and covariance matrices are easily calculated. The unconditional posterior for the coefficients is a stochastic mixture of these t distributions. Once the posterior distribution for θ is calculated, posterior simulation of the coefficients can proceed quickly.

A closed-form expression for the posterior probability distribution of θ is derived by Chin Choy and Broemling (1980):

$$p(\theta \mid \mathbf{y}) \begin{cases} \propto D(\theta)^{-n_1} \, |\mathbf{M}_1|^{-1/2}, \theta = 1, \ldots, T-1 \\ = 0, \theta = T \end{cases}, \tag{3}$$

where

$$\begin{aligned} n_1 &= n_0 + T/2 \\ D(\theta) &= s_0 + \{[\mathbf{y} - \mathbf{X}_\theta \boldsymbol{\beta}_1]'\mathbf{y} + [\boldsymbol{\beta}_0 - \boldsymbol{\beta}_1]'\mathbf{M}_0\boldsymbol{\beta}_0\}/2 \\ \boldsymbol{\beta}_1 &= \mathbf{M}_1^{-1}[\mathbf{M}_0\boldsymbol{\beta}_0 + \mathbf{X}_\theta'\mathbf{y}] \\ \mathbf{M}_1 &= \mathbf{X}_\theta'\mathbf{X}_\theta + \mathbf{M}_0 \\ \mathbf{M}_0 &= \mathbf{V}_0^{-1}s_0/n_0. \end{aligned}$$

With diffuse priors for the coefficients and precision, the posterior distribution of θ may be approximated even more simply by the likelihood in Eq. (2). We write the posterior derived by Chin Choy and Broemeling that explicitly incorporates prior information as $p_B(\theta \mid \mathbf{y})$ and the posterior based just on the likelihood function as $p_L(\theta \mid \mathbf{y})$. We compare these alternative methods for calculating θ and regression coefficients in a Monte Carlo experiment below.

Posterior probabilities of the change points, $p_B(\theta \mid \mathbf{y})$ or $p_L(\theta \mid \mathbf{y})$, can be used for Bayesian inference with stochastic sampling from the conditional posterior (SSCP). The unconditional posterior distribution of the coefficients is a mixture of t distributions located at the conditional posterior means, where the mixture probabilities are the posterior probabilities of θ,

$$p(\boldsymbol{\beta} \mid \mathbf{y}) = \sum_{\theta=1}^{T-1} t(\boldsymbol{\beta}_1, \mathbf{V}_1, 2n_1) p(\theta \mid \mathbf{y}),$$

where

$$\mathbf{V}_1 = \mathbf{M}_1^{-1} D(\theta)/n_1.$$

With diffuse priors, $\mathbf{V}_1$ is approximately equal to the OLS covariance matrix, estimated for a given value of θ. SSCP proceeds by obtaining N draws of indices, $t = 1, \ldots, T - 1$, with probability $p(\theta \mid \mathbf{y})$. The frequency of each index indicates the number of draws to obtain from each conditional posterior distribution of the coefficients. Simulating from the multivariate $t(\boldsymbol{\beta}_1, \mathbf{V}_1, 2n_1)$ begins by generating a draw from the multivariate normal distribution $N(\boldsymbol{\beta}_1, \mathbf{V}_1)$ and scaling by $w^{1/2}$, where w is a random draw from a $\chi^2/2n_1$ on $2n_1$ degrees of freedom. In contrast to the Gibbs sampler, convergence is immediate with the SSCP algorithm. Each draw from the conditional posterior distribution of the regression coefficients is independent, and the posterior distribution for θ is calculated explicitly, rather than simulated. Large numbers of iterations can be generated quickly to accurately map the shape of the unconditional posterior distribution of coefficients.

The change-point model involves a type of Bayesian model averaging (Western 1996; Bartels 1997). Bayesian model averaging provides inferences about parameters when a variety of different models, $M_1, \ldots, M_K$, may be true. For each possible model, a posterior model probability $p(M_i \mid \mathbf{y})$ is calculated. Posterior distributions of parameters are weighted sums of the posteriors under each model where the weights are given by $p(M_i \mid \mathbf{y})$. For the change-point problem, a range of models is defined by different change points. We are uncertain about the correct regression model because we are uncertain about the location of the change point and the regressors, $\mathbf{X}_\theta$. Like the usual model averaging analysis, inference about the coefficients proceeds by taking the sum of their conditional posteriors and weighting by the posterior probability of each model. In the change-point analysis, the probability of each model is given by the posterior distribution of the change point, $p(\theta \mid \mathbf{y})$.

If a likelihood function can be written for the model parameters, why introduce a Bayesian model for the change point? Maximum likelihood and Bayesian inference with a uniform prior on θ produce similar results for the location of the change point, but the methods diverge in their approaches to the regression coefficients. Maximum likelihood estimation of the coefficients conditions on the maximum likelihood estimates of θ. This approach is asymptotically justified where long time series reduce uncertainty about θ. In practice, time series are often short and considerable uncertainty accompanies the location of θ. Bayesian inference accounts for this uncertainty by integrating over, rather than conditioning on, θ. Bayesian standard errors for coefficients thus tend to be larger than the maximum likelihood estimates because the Bayesian analysis provides a more realistic accounting of prior uncertainty.

The Bayesian analysis also allows prior information about the location of a change point to be used in the analysis. In many applications, change points are hypothesized because of specific historical events that might transform a regression regime. The Bayesian model allows the location of the change point to be uncertain but also allows some changes points to be more likely than others. In this way, substantive information about a particular historical process can be used in a probabilistic way.

Large research literatures in econometrics and statistics have examined models for change points in time series. The diagnostic approach of Isaac and Griffin (1989) was statistically justified by Brown et al. (1975), who propose plots of the cumulative sum of recursive residuals—residuals obtained from out-of-sample predictions of subsets of a time series. A more elaborate parametric structure was placed on the analysis of change points by Box and Jenkins. Their transfer-function models allowed a mean shift in a response variable, where the transfer function specified the location of a change point a priori (Box et al. 1994). While transfer-function models allow more complex dynamics than often seen

in applied work, the analysis does not allow the location of a change point to be inferred from the data.

Modern time series analysis has proceeded in a variety of directions. Test statistics based on the theory of Brown et al. (1975) can be inaccurate in the presence of trended data. Hansen (2000) proposes a general test that also allows for structural change in regressors. Bai and Perron (1998) consider an alternative generalization in which a number of change points is estimated. Their analysis describes tests for the number of change points and a simple sequential method for estimation. Multiple regimes might also be specified by allowing the parameters of a model to evolve according to a time series process. These time-varying parameter models allow coefficients to change through an autoregressive process in which parameters are updated at each point in a series (Beck 1983; Harvey 1989). This model is more flexible than the Bayesian change-point model, allowing for an indeterminate number of regimes. However, the time-varying parameter models are highly parameterized and estimates can depend closely on initial values and distributional assumptions about parameters. Like Bayesian models, time-varying parameter models feature parameters that are viewed as random quantities.

Closer to the current approach that specifies the number of regimes a priori, a switching regimes model defines a response variable controlled by a Markov process. For this Markov-switching model, a transition matrix describes the probability of moving from one regime to another for each time point in a series (Hamilton 1989). The Markov-switching model is similar to the Bayesian change-point model. Both models draw a probability distribution over the space of possible regimes. Like the change-point model, structural coefficients in the switching model are a stochastic mixture where the mixture weights are given by the probability of being in a particular regime (Hamilton 1994, pp. 685–699). Bayesian analysis of the switching model, using the Gibbs sampler, is detailed by Kim and Nelson (1999). They argue for posterior simulation in a Bayesian model over maximum likelihood because the method yields finite-sample inferences and allows the inclusion of prior information.

While temporal instability in regression is an active area of research, applied researchers working with historical data in the fields of comparative politics and international relations seldom model structural changes in time series. The current Bayesian change-point model combines a simple structural specification with a coherent approach to finite sample inference. It is less rigid than the transfer-function model that specifies the change point a priori but offers more structure than the time-varying parameter model that makes no assumptions about the number of regression regimes. Like the Markov-switching models, the Bayesian change-point model takes a probabilistic approach to regime change in regression.

4 A Monte Carlo Experiment

We compare the performance of the Gibbs sampler and the SSCP based on $p_B(\theta \mid \mathbf{y})$ and based on $p_L(\theta \mid \mathbf{y})$ with a Monte Carlo experiment. In this experiment we generate $\mathbf{y}$ from

$$y_t = \beta_0 + \beta_1 x_t + \beta_3 I_t + \beta_4 I_t x_t + e_t, \quad t = 1, 2, \ldots, 30,$$

where $\beta_0 = .5$, $\beta_1 = .2$, $\beta_2 = -.3$, $\beta_4 = .2$, $I_t = 1$ if $t \geq 21$ and 0 otherwise ($\theta = 21$), x is a set of fixed regressors, and e is drawn randomly from a normal distribution with zero mean and standard deviation .5. We generate 1000 vectors, $\mathbf{y}$, and estimate the change

Table 1 Monte Carlo results for the change-point parameter, θ, estimated with the Gibbs sampler, the Bayesian closed-form expression of Chin Choy and Broemling (1980), $p_B(\theta \mid \mathbf{y})$, and posterior approximations based on the likelihood $p_L(\theta \mid \mathbf{y})$

	$E(\hat{\theta})$	$SD(\hat{\theta})$	*MSE*
Gibbs sampler	21.405	2.259	7.076
$p_B(\theta \mid \mathbf{y})$	21.417	1.977	5.915
$p_L(\theta \mid \mathbf{y})$	21.336	1.286	3.439

Note. $E(\hat{\theta})$ is the average value of $\hat{\theta}$, the point estimate of the change point in 1000 Monte Carlo trials; $SD(\hat{\theta})$ is the standard deviation of $\hat{\theta}$; MSE is the mean squared error given by the squared bias, $[E(\hat{\theta}) - \hat{\theta}]^2$, plus the variance, $V(\hat{\theta})$, where $\theta = 21$.

point and the coefficients using our three methods. To speed computation, a uniform prior is placed on θ between $t = 3$ and $t = 27$. With 2000 iterations of the Gibbs sampler, computation time was about two minutes using interpreted code written in R. Computation time with the other methods was only a second or so. (R code for the Gibbs sampler and the SSCP method is provided in the appendix.)

Monte Carlo results for θ are reported in Table 1. We examine the performance of the different methods by calculating the average value of the point estimate, $E(\hat{\theta})$. For the Gibbs sampler, $\hat{\theta}$ is the modal value of θ^* over the 2000 iterations of the Gibbs chain. For Chin Choy and Broemeling's (1980) closed-form Bayesian calculation, $\hat{\theta}$ is given by the posterior mode—that value of θ that maximizes $p_B(\theta \mid \mathbf{y})$. Finally, $\hat{\theta}$ for the method based on the likelihood function is that which maximizes $p_L(\theta \mid \mathbf{y})$ or, equivalently, $L(\theta; \mathbf{y})$. Bias for each method is given by $E(\hat{\theta} \mid \mathbf{y}) - \theta$. The dispersion of the estimates over Monte Carlo trials is measured by the standard deviation of the point estimates, $SD(\hat{\theta})$. We also report the mean squared error (MSE) given by the squared bias plus the variance of $\hat{\theta}$.

Monte Carlo results show that all three methods do similarly well, identifying the change point in the regression at $t = 21$. All three methods show a small upward bias. Estimates based on the likelihood are less dispersed than the Gibbs sampler estimates and the Bayesian closed-form calculation. This is because prior information about the precision or regression coefficients does not enter into the maximum likelihood estimation of θ. Although prior information for the Bayesian methods is intended to be diffuse, Bayesian estimates of θ integrate over uncertainty about τ and $\boldsymbol{\beta}$. Consequently mean squared error is also somewhat higher for the Bayesian methods than for the likelihood method. Estimates from the Gibbs sampler are also more dispersed than those from the closed-form calculation. This may be due to simulation error in the Gibbs chain that can be reduced by running the chain longer.

Table 2 compares the performance of the three methods for Bayesian estimation of the regression coefficients. The coefficients are estimated by their posterior expectations—averages of the simulated coefficients. In this case we report the average of the posterior expectations over 1000 Monte Carlo trials, the average of the posterior standard deviation, and the mean squared error calculated as the squared bias of the posterior expectation plus its Monte Carlo variance. Means of the posterior expectations indicate that all three methods yield approximately unbiased estimates of the regression coefficients. Biases for the Gibbs sampler and SSCP with $p_B(\theta \mid \mathbf{y})$ are slightly smaller than SSCP based on $p_L(\theta \mid \mathbf{y})$. The bias in estimates based on the likelihood approximation is more than offset by the relatively small variance. Mean squared error is consistently smaller for Bayesian estimation with $p_L(\theta \mid y)$ compared to the other methods.

Table 2 Monte Carlo results for Bayesian estimation of the regression coefficients using the Gibbs sampler and SSCP methods using $p_B(\theta \mid \mathbf{y})$ and $p_L(\theta \mid \mathbf{y})$

		Gibbs sampler	$p_B(\theta \mid y)$ *SSCP*	$p_L(\theta \mid y)$ *SSCP*
$\beta_0 = .5$	Mean	.507	.530	.534
	Mean SD	.297	.277	.260
	MSE	.088	.078	.069
$\beta_1 = .2$	Mean	.199	.211	.211
	Mean SD	.083	.074	.057
	MSE	.007	.006	.003
$\beta_2 = -.3$	Mean	−.330	−.338	−.339
	Mean SD	.496	.463	.438
	MSE	.247	.216	.193
$\beta_3 = .2$	Mean	.203	.212	.214
	Mean SD	.092	.082	.065
	MSE	.008	.007	.004

Note. Mean is the average value over 1000 Monte Carlo trials of the posterior expectation of the coefficient. Mean SD is the average posterior standard deviation. MSE is the mean squared error given by the squared bias plus the variance of the posterior expectation. SSCP is stochastic sampling from the conditional posterior.

In sum, Monte Carlo results reveal only small differences in the performance of our three methods for estimating a regression with a change point. Bias in estimates of both the change-point parameter and the regression coefficients is small, and variability in standard errors across the methods is not substantively large. Computationally, the SSCP approaches are much simpler than the Gibbs sampler. Although the SSCP methods estimate the regression coefficients with simulation, they involve only random generation from known multivariate distributions.

5 Application: Real Wage Growth in Organization for Economic Cooperation and Development Countries

Finally, we apply the Bayesian change-point model to an analysis of real data. We analyze the pooled cross-sectional time series data that are common in comparative research in political science and sociology. The challenges of modeling change points in panel data are similar to those arising in univariate series. Structural breaks in univariate series are usually motivated by an exogenous change in surrounding conditions that precipitates a change in regression regimes. With panel data, such exogenous events are thought to induce changes in the regressions for each unit in the sample. For example, Kenworthy's (2002) analysis of unemployment rates in 16 OECD (Organization for Economic Cooperation and Development) countries suggests that the impact of corporatist institutions may have declined from the 1980s to the 1990s as local-level collective bargaining increasingly determined wages. Kenworthy (2002) thus splits his panel data into two time periods, 1980–1991 and 1992–1997, to allow for the change point. This approach to panel data analysis and our change-point model both assume that the timing of a structural change is identical for all units (countries) in the sample.

Our application reanalyzes the data of Western and Healy (1999), who examined real wage growth in 18 OECD countries for the period 1965–1992. They argued that this period can be divided into two wage-setting regimes. Under the first regime, unions were able to raise wages and capture the benefits of productivity growth. Under the second

Table 3 Summary of annual percentage growth in real hourly manufacturing wage rates, 18 OECD countries

	1966–1973	*1974–1982*	*1983–1992*
Australia	2.14	1.30	−1.53
Austria	5.31	2.61	1.98
Belgium	5.91	2.90	.29
Canada	3.34	1.52	−.11
Denmark	5.75	1.70	1.15
Finland	5.55	1.17	2.19
France	5.06	3.11	.68
Germany	4.56	1.56	2.08
Ireland	6.32	3.77	.83
Italy	5.70	2.41	1.83
Japan	9.05	1.38	1.50
Netherlands	3.70	.94	.70
New Zealand	2.86	−.41	−2.05
Norway	3.79	2.00	1.73
Sweden	4.19	.22	.83
Switzerland	1.74	.67	.87
United Kingdom	3.15	1.01	2.76
United States	1.35	−.49	−.74
Average	4.38	1.45	.85

Source. Western and Healy (1999).

regime, following the first OPEC oil shock in 1973–1974, wage growth became much more sensitive to market conditions (inflation and unemployment), and the positive effects of unions and productivity growth on wages were weakened.

A summary of the wage growth data is reported in Table 3. The wage figures show the average annual growth in hourly manufacturing wage rates. In all of the 18 countries analyzed, the average annual rate of wage growth was slower in the period 1983–1992 than for 1966–1973. There is also substantial cross-national heterogeneity in average wage growth. Some countries, such as Austria and Germany, have a high general level of wage growth over the entire 1966–1992 period. Other countries, such as the United States and New Zealand, have a low general level of wage growth.

Descriptive statistics for the independent variables in the analysis are reported for each country in Table 4. Interest particularly focuses on changes in the effects of bargaining centralization, labor government, and union density. We expect that the effects of these measures of labor's power resources became weaker following the end of the golden age of postwar economic growth. For country i at time t, the shift in wage-setting regimes is modeled as

$$y_{it} = \alpha_0 + \mathbf{x}'_{it}\alpha + I_t(\theta)\beta_0 + I_t(\theta)(\mathbf{x}'_{it}\boldsymbol{\beta}) + \varepsilon_{it}, \tag{4}$$

where covariates are collected in the vector, $\mathbf{x}_{it}$, and the dummy variable $I_t(\theta)$ equals 1 for $t \geq \theta$ and 0 otherwise. The error, ε_{it}, is assumed to follow a normal distribution. We remove fixed effects from the data and reduce correlations among the predictors by subracting country-level means from y_{it} and $\mathbf{x}_{it}$ (Hsiao 1986, p. 31). The model is fit to the mean-deviated data.

Table 4 Means of the independent variables used in analysis of real wage growth in 18 OECD countries, 1966–1992

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Australia	5.03	.15	1.85	.68	.40	51.25
Austria	2.44	−.03	3.16	.33	.72	64.07
Belgium	6.93	−.06	2.80	.51	.23	70.46
Canada	7.60	−.04	1.28	.11	.67	34.57
Denmark	5.62	−.16	1.43	.78	.43	79.52
Finland	4.40	−.09	3.07	.63	.48	78.24
France	6.13	.00	2.74	.33	.31	18.60
Germany	3.68	.02	2.62	.33	.40	40.37
Ireland	7.79	.04	3.11	.54	.15	59.49
Italy	10.24	−.07	3.84	.77	.18	53.89
Japan	1.96	−.18	4.33	.33	.00	31.08
Netherlands	5.87	−.03	2.09	.63	.18	37.79
New Zealand	3.00	−.10	.90	.62	.34	40.45
Norway	2.53	−.07	2.63	.91	.56	63.34
Sweden	2.45	−.10	1.66	.85	.73	86.62
Switzerland	.51	.02	1.49	.33	.29	33.28
United Kingdom	6.68	−.04	2.01	.33	.35	50.07
United States	6.18	.05	.80	.07	.26	23.22

Note. Column headings are as follows: (1) unemployment; (2) inflation (first difference); (3) productivity growth; (4) bargaining centralization; (5) labor government; and (6) union density.
Source. Western and Healy (1999).

Exploratory analysis strongly indicates a change point in 1976. By fitting linear regressions for $\theta = 1966 \ldots, 1990$, we see that the best-fitting model is obtained at $\theta = 1976$ (Fig. 3). The R-square plot of Fig. 3 is approximately proportional to the profile log likelihood of θ and the log posterior marginal distribution under a flat prior. Because support for the change point is presented in the log scale, Fig. 3 tends to overstate evidence for structural break in the time series in nearby years. In fact, calculation in the unlogged scale shows that the posterior probability of a change point in 1976 is 92%. For these data, conventional analysis (that conditions on the change point) and Bayesian analysis (that allows prior uncertainty) yield similar results.

We compare an OLS model, assuming $\theta = 1976$, to two Bayesian models. The first puts a uniform prior on θ and constrains the error variance to be constant across countries. The second incorporates two realistic features of possible applications. Researchers analyzing panel data are often concerned about error heterogeneity across units. Beck and Katz (1995) suggest using a sandwich estimator to obtain consistent estimates of OLS standard errors (although the method is subject to finite sample biases; see Long and Ervin 2000). In Bayesian or likelihood inference, error heterogeneity is accommodated through the model specification. A simple model for the current application fits a separate error variance for each country. Each of the 18 error variances is given an inverse gamma prior distribution. We also introduce an informative prior by placing about half the probability of the time series on θ between 1968 and 1973. These dates mark the initial increase in inflation throughout the OECD area and the first OPEC oil shock. The two Bayesian models are estimated using the Gibbs sampler. Results are based on two parallel Gibbs chains of 10,000 iterations after a burn-in of 1000 iterations. Convergence diagnostics and

Fig. 3 Plot of R^2 statistics for linear regressions on wage growth data, with change points $\theta = 1966 \ldots, 1990$, 18 OECD countries 1966–1992.

inspection of the trace plots indicate adequate mixing over the parameter space. (*WinBugs* code for these models is reported in the appendix.)

Results for the regression coefficients in Table 5 contrast with OLS estimates, using the 1976 change point with the Bayesian estimates. The OLS results suggest only the significant effect of productivity growth on real wage growth from the mid-1960s to the mid-1970s. The estimate for the period effect, β_0 in the second column of Table 5, indicates that real wage growth slowed by about three percentage points across the OECD. Results for the post-1975 regime show that the effects of all predictors have moved in the negative direction. However, only the inflation effect is highly likely to have shifted from the first to the second regime. Separate analysis indicates that the bargaining level and inflation effects in the second regime are likely to be negative. Comparing the OLS results to the first Bayesian model with constant error variance suggests few differences in the results. While the Bayesian analysis tends to produce more conservative results than conventional analysis without prior information, evidence for the break point is so strong in the current data that it makes little difference if we condition on the assumption that $\theta = 1976$.

Columns 5 and 6 of Table 5 report results for the error heterogeneity model that also includes an informative prior for θ. The prior on θ places about half the prior probability from 1968 to 1973. Still the posterior expectation of θ at 1976 is almost identical to that under the noninformative prior. In the first regime, the error-heterogeneity model provides relatively strong evidence for the effects of left government, unemployment, and productivity growth. Slight improvements in the results may be due to heteroscedasticity in the data that contributes to inefficiency in OLS. In the second regime, there is strong evidence that effects of left government and inflation have become increasingly negative. The final column of Table 5 shows the net effects of the predictors in the second regime. Posterior distributions of these net effects can be calculated directly from the Gibbs output. Summing iterates of coefficients from the first and second regimes gives draws from the posterior distribution of the net effects. The net effects suggest that the coefficients of

Table 5 Coefficients (standard errors) from normal linear and Bayesian change-point models of real wage growth, 18 OECD countries, 1965–1992 ($N = 483$)

	OLS estimates		*Bayesian change point estimates*				
			Constant σ^2		*Nonconstant* σ^2		
	1966–1975 (1)	*1976–1992 (2)*	*First regime (3)*	*Second regime (4)*	*First regime (5)*	*Second regime (6)*	*Second regime net effects (5)+(6)*
Change point (θ)	-	1976.00	-	1976.03 (.19)	-	1976.02 (.19)	-
Intercept	1.80 (.31)	−2.73 (.36)	1.78 (.32)	−2.72 (.11)	1.43 (.32)	−2.25 (.28)	−.82 (.15)
Bargaining level	−.41 (.78)	−.82 (.96)	−.41 (.78)	−.82 (.96)	.18 (.78)	−1.27 (.96)	−1.09 (.72)
Left government	.57 (.50)	−.88 (.62)	.57 (.51)	−.88 (.63)	.85 (.46)	−1.31 (.63)	−.46 (.33)
Unemployment	−.07 (.11)	−.17 (.13)	−.07 (.11)	−.17 (.13)	−.15 (.10)	−.16 (.13)	−.30 (.06)
Union density	.08 (.21)	−.55 (.31)	.07 (.21)	−.54 (.32)	.01 (.19)	−.44 (.32)	−.43 (.20)
Inflation	.03 (.06)	−.22 (.08)	.03 (.06)	−.22 (.09)	.02 (.06)	−.27 (.09)	−.25 (.06)
Productivity growth	.22 (.08)	−.11 (.11)	.22 (.08)	−.10 (.11)	.26 (.08)	−.16 (.11)	.10 (.07)

Note. Coefficients for the intercept are estimates of α_0 and β_0 in Eq. (2). Union density coefficients have been multiplied by 10. The nonconstant error variance model also includes an informative prior θ. The change point of 1976 for the OLS model is set by assumption.

unemployment, union density, and inflation became large and negative after the mid-1970s. However, we cannot confidently conclude that wages rise with productivity as they did in the late 1960s.

6 Discussion

Although political scientists and sociologists have become sensitized to the possibility of historical variability in quantitative results, conventional methods for studying that variability have several important shortcomings. Commonly, a putative change point in a time series is chosen a priori, and this is modeled with a dummy variable or else the sample is split into two periods. In this approach, the location of the change point is simply assumed and an analysis may yield positive results even though the data may strongly indicate a change point at a different time. Alternatively, Isaac and Griffin (1989), in an important paper on temporal instability in regression, suggest a diagnostic approach that allows the location of a change point to be found empirically. Although their approach has some advantages over assuming a change point's location, the diagnostic method has not been widely used because it fails to provide statistical inferences about structural change.

We describe a Bayesian change-point analysis of structural change in a time series. The Bayesian analysis, like Isaac and Griffin's (1989) method, helps diagnose the location of structural change in regression relationships. But like standard parametric approaches to period effects, the Bayesian analysis also provides statistical inferences about regression coefficients. Improving over both the usual approaches, Bayesian analysis also provides a statistical inference about the location of the change point. Indeed, the location of the change point can be simply found with standard statistical software. A Monte Carlo experiment showed that a Bayesian change-point model is feasibly estimated with the Gibbs sampler or by stochastically sampling from the conditional posteriors for the regression coefficients.

The Bayesian model was illustrated in an analysis of real wage growth in the OECD countries between 1966 and 1992. Although many researchers claimed that this recent period of slow economic growth marks the end of the golden age of postwar economic expansion, the idea of two wage regimes in the postwar political economy has not been rigorously examined. The Bayesian analysis provides clear evidence of a structural break in the wage growth process in 1976. This analysis indicates the clear importance of a sensitivity to historical variation in large-scale causal processes. We believe the analysis also underlines the importance of methods that can detect and provide statistical inferences about such variability.

Appendix

R Code for the Gibbs Sampler and SSCP

```
#  bcp.R
#  R code for Bayesian change point estimation described in the paper,
#  Bruce Western and meredith Kleykamp, "A Bayesian Change Point Model for
#  Historical Time Series Analysis." The paper describes a Gibbs sampler
#  and several approaches to stochastic sampling from the conditional
#  posterior (SSCP) of regression coefficients.
#  Bruce Western, April 27, 2004
```

```
library(MASS)
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
#   Defining some useful functions
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
#   Linear predictor function
lp <- function(x,b) as.vector(apply(t(x)*b,2,sum))

#   Normal likelihood function
n.like <- function(eta,y,sigma) {
ll <- sum(dnorm(y,mean=eta,sigma,log=TRUE))
exp(ll) }

#   Determinant of a matrix
det <- function(x) prod(eigen(x)$values)

#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
#   Gibbs function
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
gibbs <- function(X, y, tstar=length(y)/2, iter=500, burn=100) {
#   User supplies X, a matrix including unit vector in col. 1 for intercept
#                  y, a dep var vector
#                  tstar, start value for the change-point
#                  iter, number of iteration in the gibbs sampler
#                  burn, number of burn-in iterations

#   OLS fit
out <- lm(y~X-1)
p <- ncol(X)
n <- length(y)

#   Setting up data
grid1 <- (p+1):(n-p-1)

XX <- as.list(grid1)
for(i in grid1) {
tt <- (1:n)>=i
XX[[i-min(grid1)+1]] <- cbind(X, X * tt)
}

#   Priors
pb <- rep(0,2*p)
pvb <- diag(rep(1e6,2*p))
pvbi <- diag(1/rep(1e6,2*p))
ps2 <- .001
pdf <- .001

#   Start values
bstar <- rep(out$coef,2)

#   Initializing parameter vectors
beta <- h <- tt <- NULL
```

```
k <- length(grid1)
temp <- rep(NA,k)
thetas <- grid1

#   Posterior variance quantities
posta <- (n/2)+pdf

for(i in 1:(iter+burn)) {
indt <- as.numeric((1:n)>=tstar)
XZ <- cbind(X, X*indt)
xx <- crossprod(XZ)
xy <- crossprod(XZ,y)
yhat <- lp(XZ, bstar)
estar <- y-yhat
posts2 <- ps2+sum(estar^2)
hstar <- rgamma(1,posta,rate=posts2/2)
postvb <- solve(pvbi+(xx*hstar))
postb <- as.vector(postvb %*% (xy*hstar))
bstar <- as.vector(mvrnorm(1,postb,postvb))
temp <- unlist(lapply(lapply(XX,lp,bstar),n.like,y,sqrt(1/hstar)))
postt <- temp/sum(temp)
tstar <- sample(thetas, 1, prob=postt)
if(i>burn){
beta <- rbind(beta,bstar)
h <- c(h,hstar)
tt <- c(tt,tstar)
} }
mode <- table(tt)
mode <- names(mode)[mode==max(mode)]
output <- list(grid1,tt,h,beta,as.numeric(mode))
names(output) <- c("time","theta","precision","beta","mode of theta")
output
}

#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
#   SSCP function
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

sscp <- function(X, y, pmu, pvb, pa, pb, bayes=T) {
#   User supplies: X, a matrix including unit vector in col. 1 for intercept
#                  y, a dep var vector
#                  pmu, vector of prior means
#                  pvb, prior covariance matrix for coefs
#                  pa, prior degrees of freedom
#                  pb, prior sums of squares
#                  bayes=T, coefs based on bayes posterior probs
#                  bayes=F, coefs based on llik posterior probs

n <- length(y)
p <- ncol(X)
```

```
#   Grid over which break point is searched
ind <- (p+1):(n-p-1)

tau <- solve(pvb)*pb/pa

betam <- betapv <- ppr <- as.list(ind)
ppm <- ll <- rep(NA,length(ind))
postb <- NULL
astar <- pa + (n/2)

for(i in ind) {
indt <- as.numeric((1:n) >= i)
xm <- cbind(X, indt*X)
ols <- lm(y ~ xm - 1)
yhat <- predict(ols)
sigma <- summary(ols)$sigma
ll[i-p] <- n.like(yhat, y, sigma)
pxpx <- crossprod(xm) + tau
betam[[i-p]] <- solve(pxpx) %*% (tau%*%pmu + crossprod(xm, y))
e <- as.vector(y - (xm%*%betam[[i-p]]))
b2 <- as.vector(crossprod(pmu - betam[[i-]],tau%*%pmu))
dm <- pb + sum(e*y)/2 + b2/2
ppr[[i-p]] <- (astar/dm) * pxpx
ppm[i-p] <- dm^(-astar) * det(pxpx)^(-.5)
}
ppm <- ppm/sum(ppm)
ppll <- ll/sum(ll)
output <- cbind(ppm,ppll)
dimnames(output) <- list(as.character(ind),c("Bayes","LL"))

#   SSCP for coefficients
if(bayes) {
mixind <- table(sample(ind,10000,replace=TRUE,prob=ppm))
mode <- ind[ppm==max(ppm)] }
else {
mixind <- table(sample(ind,10000,replace=TRUE,prob=ppll))
mode <- ind[ppll==max(ppll)] }
newind <- as.numeric(names(mixind))
for(i in 1:length(newind)) {
mu <- as.vector(betam[[newind[i]-p]])
Sigma <- solve(ppr[[newind[i]-p]])
normb <- mvrnorm(mixind[i], mu, Sigma)
tb <- normb/sqrt(rchisq(1,astar)/astar)
postb <- rbind(postb,tb) }

output <- list(ind, output, postb, as.numeric(mode))
names(output) <- c("time","p(theta | y)","beta","mode of theta")
output
}
```

Winbugs Code for the Analysis of OECD Data

```
#   Model 1, constant error variance
model
{
for(i in 1:N) {
   y[i] ~ dnorm(mu[i], tau)
   mu[i] <- alpha0 + inprod(alpha[],x[i,]) + beta0*J[i] + J[i]*inprod(beta[],x[i,])
   J[i] <- step(yr[i] - cp - 0.5)
}
for(i in 1:p) {
   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)
cp ~ dcat(priort[])
tau ~ dgamma(.001, 0.001)
}


#   Model 2, heterogeneous error variance
model
{
for(i in 1:N) {
   y[i] ~ dnorm(mu[i], tau[cc[i]])
   mu[i] <- alpha0 + inprod(alpha[],x[i,]) + beta0*J[i] + J[i]*inprod(beta[],x[i,])
   J[i] <- step(yr[i] - cp - 0.5)
}
for(i in 1:p) {
alpha[i] ~ dnorm(0.0, 1.0E-6)
beta[i] ~ dnorm(0.0, 1.0E-6)
}
for(i in 1:NC) {
   tau[i] ~ dgamma(.001, 0.001)
}
for(i in 1:T) {
   priort[i] <- prior[i]/sum(prior[])
}
alpha0 ~ dnorm(0.0, 1.0E-6)
beta0 ~ dnorm(0.0, 1.0E-6)
cp ~ dcat(priort[])
}
```

References

Bai, Jushan, and Pierre Perron. 1998. "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica* 66:47–78.

Bartels, Larry M. 1997. "Specification Uncertainty and Model Averaging." *American Journal of Political Science* 41:641–674.

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

Beck, Nathaniel, and Jonathan Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data." *American Political Science Review* 89:634–647.

Box, George E. P., Gwilym M. Jenkins, and Gregory C. Reinsel. 1994. *Time Series Analysis: Forecasting and Control,* 3rd ed. Englewood Cliffs, NJ: Prentice Hall.

Brown, R. L., J. Durbin, and J. M. Evans. 1975. "Techniques for Testing the Constancy of Regression Relationships over Time" [with discussion]. *Journal of the Royal Statistical Society, Series B* 37:149–192.

Büthe, Tim. 2002. "Taking Temporality Seriously: Modeling History and the Use of Narrative as Evidence." *American Political Science Review* 96:481–494.

Carlin, Bradley P., and Thomas A. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis.* New York: Chapman and Hall.

Chin Choy, J. H., and L. D. Broemeling. 1980. "Some Bayesian Inferences for a Changing Linear Model." *Technometrics* 22:71–78.

Freedman, David A. 1983. "A Note on Screening Regression Equations." *American Statistician* 37:152–155.

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57:357–384.

Hamilton, James D. 1994. *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hansen, Bruce E. 2000. "Testing for Structural Change in Conditional Models." *Journal of Econometrics* 97: 93–115.

Harvey, Andrew C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge, UK: Cambridge University Press.

Hsiao, Cheng. 1986. *Analysis of Panel Data.* Cambridge, UK: Cambridge University Press.

Isaac, Larry, and Larry J. Griffin. 1989. "Ahistoricism in Time-Series Analyses of Historical Process: Critique, Redirection, and Illustrations from U.S. Labor History." *American Sociological Review* 54:873–890.

Jackman, Simon. 2000. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo." *American Journal of Political Science* 44:375–404.

Kenworthy, Lane. 2002. "Corporatism and Unemployment in the 1980s and 1990s." *American Sociological Review* 67:367–388.

Kim, Chang-Jin, and Charles R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Application.* Cambridge, MA: MIT Press.

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

Long, J. Scott, and Laurie H. Ervin. 2000. "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model." *American Statistician* 54:217–224.

Mitchell, Sara McLaughlin, Scott Gates, and Havard Hegre. 1999. "Evolution in Democracy-War Dynamics." *Journal of Conflict Resolution* 43:771–792.

Richards, Mark J., and Herbert M. Kritzer. 2002. "Jurisprudential Regimes in Supreme Court Decision Making." *American Political Science Review* 96:305–320.

Schickler, Eric, and Donald P. Green. 1997. "The Stability of Party Identification in Western Democracies: Results from Eight Panel Surveys." *Comparative Political Studies* 30:450–483.

Thelen, Kathleen, and Sven Steinmo. 1992. "Historical Institutionalism in Comparative Politics." In *Structuring Politics: Historical Institutionalism in Comparative Perspective*, eds. Sven Steinmo, Kathleen Thelen, and Frank Longstreth. New York: Cambridge University Press, pp. 1–32.

Western, Bruce. 1996. "Vague Theory and Model Uncertainty in Macrosociology." *Sociological Methodology* 26:165–192.

Western, Bruce, and Kieran Healy. 1999. "Explaining the OECD Wage Slowdown Recession or Labour Decline?" *European Sociological Review* 15:233–249.