Clustering time: applying Bayesian mixture models to estimate temporally heterogeneous effects in longitudinal analysis

Paasha Mahdavi
Department of Political Science
University of California, Los Angeles
paasha@ucla.edu

September 6, 2013

Prepared for delivery at the 2013 Annual Meeting of the
American Political Science Association, August 29-September 1, 2013.

Abstract

The presence of temporally heterogeneous effects is prevalent in longitudinal (panel) analysis in the social sciences. The effects of predictors on outcomes of interest may vary across time, often following complex patterns. Though unit heterogeneity is more commonly addressed in quantitative studies using longitudinal data, a growing body of literature has begun to directly model the presence of time-varying effects using methods such as time fixed-effects, time interactions, unstructured time models, structural break models, and dynamic linear models. This study considers an alternative approach that allows researchers to answer questions regarding (1) temporally heterogeneous effects and (2) how these effects are clustered over time. Using the debate surrounding the presence of a resource curse (Ross 2012) as an example, we apply a Bayesian mixture modeling (BMM) framework to address time-varying effects of oil wealth on democratic governance. Results indicate that the BMM approach provides evidence for the presence of a resource curse for the periods 1960-1987 and 1995-2003, with null effects in the 1987-1990 period and positive effects in the 1991-1994 era. The advantage to the BMM framework, we argue, is the lack of *ad hoc* temporal modeling which can often lead to high model dependence; instead by using a data-based approach to temporal clustering, we flexibly allow for hypotheses of theoretical interest to be tested against the data rather than be assumed by the model.

Keywords: Longitudinal analysis, temporal heterogeneity, dynamic effects, clusters, Bayes, mixture models, resource curse

1 Introduction

The presence of temporally heterogeneous effects is prevalent in longitudinal or panel analysis in political science as the effects of predictors on outcomes of interest often vary across time. Scholars of comparative politics and international relations generally have strong theoretical priors about the dynamic nature of political forces, yet current studies typically only report average effects across time.

Though unit heterogeneity is more commonly addressed in quantitative studies using longitudinal data, a growing body of literature has begun to directly model the presence of time-varying effects using time fixed-effects, time interactions, unstructured time models, structural break models, and dynamic linear models. This paper considers a data-based approach to the issue of temporally heterogeneous effects. Using the well-established framework of Bayesian mixture modeling, we show that clustering random effects based on time addresses the issue of temporally heterogeneous effects in a completely data-based (as opposed to *ad hoc*) approach in order to test hypotheses of theoretical interest related to how effects change over time.

Within the political economy of development literature, a growing debate has emerged on whether or not oil and natural resources hinder democracy and instead promote authoritarian governance (Aslaksen 2010; Haber and Menaldo 2011; Ross 2001, 2012; Ulfelder 2007). A central critique of the early arguments in favor of the "resource curse", as the theory has come to be known, is that the pernicious effects of oil on democracy are limited to a small window of time in the 1960-70s and early 2000s when oil prices were high and states began nationalizing their oil sectors (Haber and Menaldo 2011; Musgrave and Liou 2010). In other time periods, the negative effect of natural resources on regime type disappears (Haber and Menaldo 2011) or, in some cases, is positive (Dunning 2008).

This example illustrates the possibility of temporally heterogeneous effects in a way that is not easily parameterized. To show this pattern visually, we replicate findings from Ross (2012) and then apply an unstructured time model where year fixed-effects are interacted with the oil variable.[1] The results are plotted in Figure 1, which shows the mean coefficients with standard errors of the effect of oil wealth on polity scores across time.

[1] Andersen and Ross (2013) employ a similar method to examine the temporal heterogeneity of the effects of oil fiscal reliance on polity scores. Specifically, see Figure 6, page 17.

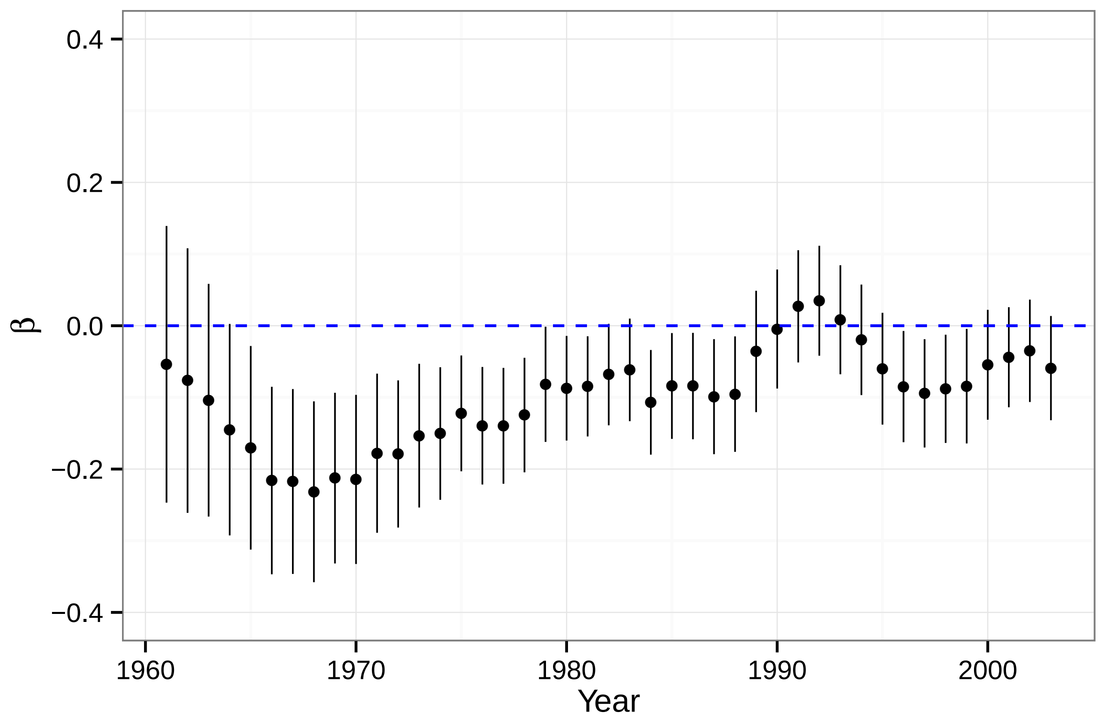

Figure 1: Coefficient estimates from an unstructured time model (see text for definition of this model) linear regression of oil income per capita on polity scores, controlling for lagged polity scores (akin to an AR(5) process) and logged income per capita. See Ross (2012), Table 3.8.

What makes this example particularly representative of the temporally heterogeneous effects problem is the complex pattern of the effect estimates over time. If instead the graph showed a monotonic increase in effect size over time, then a simple solution would be a time-effect lower order polynomial interaction. Or if there were a clear break point before and after which the effect size differed, then applying a structural break or change point model would appropriately address the problem. However, there are numerous instances in political economy in which effects vary in more complex ways so that they cannot be easily parametrized. The study of civil war onset and duration, democratization, interstate alliance formation, comparative distributive politics (especially public goods provisions over time), and the determinants of economic growth are just a few topics for which temporal heterogeneity of effects is likely to exist.

To address the problem of temporally heterogeneous effects we propose a Bayesian mixture modeling approach. This method builds on existing modeling solutions to temporal heterogeneity, notably work on structural breaks and Bayesian change point analysis. The key difference in the

Bayesian mixture modeling approach is that we can explicitly model temporal heterogeneity in effects *without* assuming temporal heterogeneity in the outcome and other predictors of interest. In the next sections, we outline current solutions to this problem and explain the Bayesian mixture model approach and how it differs. Following this discussion, we apply the framework to the example shown above from Ross (2012) and compare our results to existing modeling solutions. We conclude with a discussion of the implications of this approach along with potential avenues of future research.

2 Current modeling solutions

In this section, we review current approaches to address and estimate temporal heterogeneity in longitudinal data. We move from the simplest approach (standard linear regression with no time modeling) to a more structured approach (structural break and change point modeling) before discussing our Bayesian mixture modeling framework.

2.1 Standard linear regression

Building from first principles, consider a model that pools the effects of a predictor of interest X on an outcome variable Y using panel data with units i and time points t. This would be specified with the conventional linear regression framework, written as:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \epsilon_{it}, \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{1}$$

where we include a fixed intercept β_0 and a random error term, typically assumed to follow a standard normal distribution. In this context, β_1 is often interpreted as the effect of X on Y, on average across time periods t.

2.2 Time fixed-effects and interactions; Unstructured time models

If we have reason to believe that the effects of a predictor of interest changes over time, we can use time fixed-effects with interactions in the standard linear regression framework. The simplest approach, often called the "*ad hoc* method for studying change points", is to add a dummy variable I_t, which is set to 0 for all years prior to a specified year and 1 thereafter, and apply an interaction with the predictor of interest X_t (Western and Kleykamp 2004). This is modeled as such:

$$Y_{it} = \beta_0 + \beta_1 I_t + \beta_2 X_{it} + \beta_3 I_t X_{it} + \epsilon_{it} \tag{2}$$

In this approach, the researcher must specify the time point for the measurement of I_t, typically providing some theoretical justification for this choice. Often citing Waltz's seminal work on structural realism after the Cold War, scholars apply a "Cold War dummy" to regressions on topics ranging from great power alliances, UN interventions, and interstate conflict (Sweeney and Fritz 2004; Beardsley and Schmidt 2012; Chiozza 2002). In these examples, the researcher adds a dummy variable equal to 1 prior to 1989 and 0 thereafter (or the reverse if used as a "post-Cold War dummy") and applies an interaction to the predictor(s) of interest. Typically using a standard t-test, the hypothesis of temporal heterogeneity of effects pre- and post-Cold War is rejected if the coefficients for the predictor and interaction terms are not statistically significant at conventional levels. In this case, there is a strong and clear theoretical motivation for choosing 1989 as a break point; this may not generally be the case with longitudinal topics dealing with dynamic effects.

In addition, several different choices of years for I_t can be separately modeled in successive regressions, with the "best" time point selected by comparing the R^2 diagnostic term across all models. Using the optimal time point, for example I_t^*, the model is then run with I_t^* and coefficient estimates are obtained via ordinary least squares. While this approach is statistically sound in that it generates unbiased estimates, the downside is succinctly pointed out by Western and Kleykamp (2004, 357):

> After estimating $\hat{\theta}$ ... we can calculate the regression coefficients, their standard errors, and p values using OLS. Conventional standard errors and p values will be too small, however, increasing the risk of an incorrect inference of statistically significant effects. ... By using 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.

An adjustment to this approach is the unstructured time model, where we remove the dummy and predictor terms and only consider the interaction term. Instead of specifying the time for the dummy term, instead we apply the full range of time dummies as interactions. Note that this is similar to running T cross-sectional regressions, though by doing so we are implicitly assuming independence across each cross-sectional regression given the estimation of a new ϵ_i for each t. Instead, the unstructured time model estimates one "grand mean" error term in ϵ_{it}. Thus the unstructured time model can be written as such:

$$Y_{it} = \beta_0 + \sum_{t=1}^{T} \beta_t I_t X_{it} + \epsilon_{it}, \qquad t = 1, \ldots, T \tag{3}$$

where the interactions of the predictor of interest and a time dummy are summed for each time period t in the sample. Overall, the advantage to the time fixed-effects and unstructured time models is that they provide a quick check for time-varying effects. In particular, the unstructured time model allows the researcher to visualize temporal heterogeneity by constructing bivariate plots as we do in Figure 1. While this modeling framework allows the researcher to identify the heterogeneity of effects across individual time points (in this case, years), the disadvantage is that there is no clear-cut way to group time-varying effects into meaningful clusters – instead, the unstructured time model can be considered to provide T separate clusters, one for every time point in the data.

2.3 Dynamic linear models

A more flexible modeling framework for longitudinal data is the dynamic linear model (DLM). The dynamic modeling approach treats the predictor coefficient β_t as a time-varying parameter to be estimated using a structural equations framework (Martin and Quinn 2002; West and Harrison 1997). In this fashion, each successive β_t after $\beta_{t=1}$ is modeled as a function of the prior predictor coefficient β_{t-1}. The general multivariate DLM is specified as follows:

$$Y_{it} = \beta_0 + \mathbf{X'}_{it}\beta_t + \epsilon_{it} \tag{4}$$

$$\beta_t = \mathbf{W}_t\beta_{t-1} + \delta_t \tag{5}$$

where $\mathbf{X}_{it}$ is a matrix of P covariates with $T \times N$ rows, $\mathbf{W_t}$ is a $P \times P$ matrix of covariates at time t, and ϵ_{it} and δ_t are stochastic error terms with i.i.d. normal distributions. The predictor coefficient is explicitly modeled in a stochastic framework and estimated simultaneously with the regression model for the outcome variable Y. The second equation is specified such that each estimate of β_t follows a relatively smooth pattern over time. In the Bayesian framework, the DLM can be thought of as a likelihood function plus a prior distribution on the random variable β_t.

Though the DLM is a more flexible framing of temporal heterogeneity than the previous models, the constraint here is that β_t is forced to depend on previous β's as in a Markov process (so that each successive estimate of β_t will be correlated with the previous β_{t-1}). This may not necessarily be the case for longitudinal processes in the social sciences, where the effects of given predictors vary by year in a non-sequential fashion. Consider the literature on the determinants of civil wars where a strong predictor of civil war onset is state capacity (Fearon and Laitin 2003). The negative effect of state capacity on conflict onset may be reduced in years with exogenous economic shocks such as the 1973 and 1979 oil price shocks or the 2001 terrorist attacks on the United States, because in these years conflict onset may be more likely despite strong state capacity for economic grievances not attributable to the state. Because a strong state can recover quickly after an economic shock (depending on the magnitude of the shock), the effect of state capacity on war onset may be smaller only in these years but not the "contiguous" years of each shock (e.g. 1980 for the oil price shock of 1979). In these cases, a more flexible framework allowing structurally different parameter estimates for successive time points is preferred.

2.4 Structural breaks and change points

A more non-parametric approach is found in the analysis of change points or structural breaks. Here, temporal variation in univariate or multivariate longitudinal models is captured by estimating the location of breaks at certain points in time (Park 2011; Spirling 2007). Building on the model specified in (2) the standard change point model is specified as follows:

$$Y_{it} = \beta_0 + \beta_1 I_t(\kappa) + \beta_2 X_{it} + \beta_3 I_t(\kappa) X_{it} + \epsilon_{it} \tag{6}$$

where $I_t(\kappa) = 0$ for $t < \kappa$ and $I_t(\kappa) = 1$ for $t \geq \kappa$. This allows us to treat the change point as a parameter to be estimated while also specifying prior distributions for the model parameters (Western and Kleykamp 2004). We can rewrite the model parameters using matrix notation as $X_{it\kappa}$, which is a matrix with columns X_{it}, $I_t(\kappa)$, and $I_t(\kappa)X_{it}$. Assuming that y is normally distributed conditional on model parameters, we can write out the likelihood and prior distributions:

$$Y_{it}|\beta, \mathbf{X_{it\kappa}} \sim N(\mathbf{X'_{it\kappa}}\beta, \tau^{-1}) \tag{7}$$

$$\beta \sim N(m, v) \tag{8}$$

$$\kappa \propto (T-1)^{-1} \tag{9}$$

$$\tau \sim \text{Gamma}(a, b) \tag{10}$$

where m, v are prior means and variances for β, and a, b are hyperpriors for the precision τ. Note that the prior distribution for the change point parameter κ is uniform, thereby setting equal prior probability to each change point possibility.[2]

Park (2012) applies a different approach to structural breaks to address dynamic heterogeneity. Instead of estimating the number of change points parametrically, Park suggests running K structural break models and employing Bayesian model selection to identify the appropriate number of breaks (according to each model's Bayes factor). In this framework the dynamic heterogeneity is modeled in terms of the regression function, where breaks in how $Y|X, \beta$ changes over time are modeled with the optimal number of change points, not decided by the researcher but instead decided by the data (via Bayesian model selection).

The model framework for the structural breaks approach is mathematically similar to the Bayesian mixture model framework we suggest in this paper, though our model differs only slightly from this framework. Whereas the structural breaks method explicitly models breaks in the likelihood function (or outcome variable, for univariate analysis), it does not consider breaks only in the parameters without any break in the likelihood. That is, the structural breaks approach only considers dynamic heterogeneity in the regression model as a whole, without allowing breaks only for dynamic heterogeneity in effects (with dynamic homogeneity in the likelihood function). For example, given the likelihood function $p(Y|\theta)$, though the longitudinal profile of the outcome variable Y does not exhibit structural breaks, the underlying parameters θ could show patterns of structural breaks across time. Thus, applying a model that treats a structural break as a parameter to be

[2]In the non-Bayesian framework, the structural break is typically estimated by using a Chow test for every potential break point. This is done by fitting an OLS model for the observations before and after the potential change point, and computing the error sum of squares. To get the restricted sum of squares, another OLS model for all observations is computed. See Andrews (1993).

estimated with respect to the entire likelihood function – as is the case in Bayesian change point analysis – may be considered "overkill" if the only structural break we believe to exist is found in how certain effects change over time.

Further, the structural breaks approach can be considered a more restricted type of finite mixture model. Instead of allowing individual time points to be clustered freely, the change point model restricts time points to be assigned to the same cluster before and after a given structural break. For example, for data with $T = 25$ time points, if the change point model identifies breaks at $t = 5$ and $t = 12$, then all time points where $t < 5$ are forced into one cluster, all time points such that $5 \leq t < 12$ are in one cluster, and all $t \geq 12$ are in a third cluster. This approach might be preferred for some substantive questions where time-varying effects are slow-changing, such as the effects of income inequality on democracy (Acemoglu and Robinson 2005) or geographic causes of economic development (Sachs 2003). However, by considering a more flexible approach the hypothesis of slow-varying dynamic effects can be directly tested instead of being taken as a model assumption. The Bayesian mixture model framework, as we argue in more detail below, is one such flexible approach to clustering time points for processes with temporally heterogeneous effects.

3 The Bayesian mixture model framework

Building on the structural break and change point approach is the proposed Bayesian mixture model framework. The intuition behind the mixture model is that there is some unobserved clustering of parameters across some unknown number of groups, denoted as k clusters or mixtures. Each group k follows its own distribution with its own mean and variance (or other distributional parameters, for the non-Gaussian case). For our case, the coefficient of the temporally heterogeneous effect is the parameter being mixed across k clusters of time points t. Our inferential goal, therefore, is the posterior distribution of β_t given the clustering assignment and the data. That is, we can directly answer questions of how effects are clustered across time.

In the case of the linear mixed model for longitudinal data, we typically have a repeated-measure response variable that depends on known covariates $\mathbf{X}$ and unknown regression coefficients, which include a vector of fixed effects α and time-specific random effects which we can denote β_t.[3] To

[3]More often it is the case to have unit-specific random effects β_i to account for spatial heterogeneity either using random intercepts or random slopes for select coefficients.

be explicit, $\mathbf{X}$ can be partitioned into a matrix of covariates $\mathbf{X}_{it}$ to be estimated with fixed effects coefficients α and a matrix of covariates $\mathbf{W}_t$ estimated with time-specific random coefficients β_t. In addition, to represent the clustering assignment we have the latent variable $z_t \in \{1, \dots, K\}$ for $t = 1, \dots, T$. Here, $z_t = k$ when the tth time random effect vector is sampled from the kth population cluster. The probability that the time random effect vector β_t is assigned to group k is given by the conditional probability $P(z_t = k|\mathbf{w}) = w_k$, where $\mathbf{w}$ is a vector of the individual cluster weighting probabilities w_k. Given a cluster assignment, the posterior distribution of the random effect parameter β_t is weighted distribution of k normals with means μ_k and covariance matrix D_k. Using this notation, and keeping consistent with the above modeling frameworks, we can write the Bayesian linear mixture model as follows:

$$
\begin{aligned}
Y_{it}|\alpha, \beta_t, \mathbf{X_{it}}, \mathbf{W}_t, \tau &= \mathrm{N}(\mathbf{X'}_{it}\alpha + \mathbf{W'}_t\beta_t, \tau^{-1}) && (11)\\
\alpha &\sim \mathrm{N}(m, v) && (12)\\
\beta_t|\mu_k, \mathbf{D}_k, z_t &= \sum_{k=1}^{K} P(z_t = k|\mathbf{w}) p(\beta_t|z_t = k, \mu_k, \mathbf{D}_k), \qquad t = 1, \dots, T && (13)\\
&= \sum_{k=1}^{K} w_k * \mathrm{N}(\beta_t; \mu_k, \mathbf{D}_k), \qquad k = 1, \dots, K && (14)\\
\tau &\sim \mathrm{Gamma}(a, b) && (15)
\end{aligned}
$$

where the difference from the structural breaks approach is the introduction of the latent variable z_t for cluster assignment and the weighted distribution of a random effects parameter β_t. In our case, this β_t is the parameter of interest – it represents the coefficient vector for the researcher's predictor of interest, which is allowed to vary over time. The researcher supplies the priors $a, b, m, v, \mu_k, \mathbf{D}_k$, though these can be set as quasi-informative vague priors. With this framework in mind, we have one additional step to specify: the prior distribution of the mixture parameter z_t. The literature on mixture modeling typically assumes that z follows a Multinomial distribution with parameters λ, which follow a Dirichlet distribution with equal probabilities $\alpha_1 = \cdots = \alpha_K$ (Box and Tiao 1973; Dempster, Laird and Rubin 1977; Gelman et al. 2004). We can write this out as follows:

$$\mathbf{z}_t|\lambda \sim \text{Multinomial}(1, (\lambda_1, \ldots, \lambda_K)) \tag{16}$$

$$\lambda \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_K) \tag{17}$$

This specification is generally estimated using priors that assign equal weight to clustering assignment. That is, *a priori* the probability of a time point t being assigned to any cluster $k \in (1, \ldots, K)$ is $\frac{1}{K}$ for all clusters. Further, in the simplest specification of the mixture model there is no assumption of time dependence: the probability for a given time point t to be assigned to cluster k does not depend on the cluster assignment of time points $t-1$ or $t+1$. Time dependence can be built in to the model by setting the covariance matrix of the random effects mixture $\mathbf{D}_k$ to some covariance structure that represents temporal dependence (e.g. AR(p), MA(q), ARMA(p,q), Equicovariance, etc.).

To address temporal heterogeneity of a given effect, we can set priors for different cluster means based on the research question at hand. For example, if we are testing to see if the effect of X on Y is positive for some time periods, negative for others, and zero for others, we can set the number of clusters $K = 3$ with prior means $\mu_1 < 0$, $\mu_2 = 0$, and $\mu_3 > 0$ with relatively small variance. In this case, by looking at the posterior cluster probabilities and assignments, we can see how many time points t were assigned to each group. This can be helpful in testing hypotheses of negative vs. null vs. positive effects, as our example in the next section will show. In other words, by allowing a data-based clustering of time points into groups, we can not only discern the temporal heterogeneity of effects (or homogeneity, if all time points fall into one cluster), but we can also observe the frequency of group assignment. If for example, most time points fall into the "null" cluster, this provides some support to claims that the effect of X on Y is mostly null across time. Still, this should not be considered as an *ad hoc* test of substantive hypotheses. Rather, the mixture model approach should be augmenting theoretical implications of temporal heterogeneity.

Though it is somewhat cumbersome to fully specify the Bayesian mixture model, the advantages of using this approach are four-fold. First, we are able to estimate mixtures of parameters of interest without having to allow mixtures of the entire likelihood function (as is the case in the structural breaks approach). Second, we are writing out nearly every step (except hyperpriors) of the model so that nothing is left "under the hood". Third, a key advantage of the mixture modeling approach is that when the computation of the model fails (i.e. does not reach appropriate convergence), the model assumptions have been violated (Gelman et al. 2004). This is an advantage because it provides some indication to the researcher that the proposed model is inappropriate given the data, which is a diagnostic tool not available in the previously discussed current modeling solutions to temporal heterogeneity. Finally, a fourth advantage of this approach is that if indeed there is no temporal heterogeneity of effects, then the posterior estimates will suggest only *one* cluster instead of $k > 1$ clusters.[4]

4 Application to Ross (2012)

With these different approaches to addressing temporally heterogeneous effects in mind, we now apply the Bayesian mixture model to Ross's (2012) oil-democracy example from the introduction. To begin, we must state the model assumptions up front. (1) We assume quasi-informative priors for all fixed effects parameters ($m = 0, v = 10$) and the stochastic parameter ϵ (Gamma parameters $a = 0.2, b = 0.1$). (2) We assume the random effects parameter which we will mix – in this case parameter is the coefficient of logged oil income – follows a mixture of normal distributions. In our case, we are testing the hypothesis of temporal heterogeneity of effects with three substantively interesting possibilities: negative effects of oil on democracy (Ross 2012), null effects (Haber and Menaldo 2011), and positive effects (Dunning 2008). Thus, we set prior mixture means of -0.1, 0.0, and 0.1 with tight variances (0.01).[5] (3) We assume the three mixture weights (w_1, w_2, w_3) are set equal prior probability at $\frac{1}{3}$.

With this framework, we are particularly interested in the cluster assignments: if Ross's hypothesis of a negative effect of oil on democracy is true, then we will observe a high number of time points (years) t assigned to cluster 1. If otherwise – that is, if there is no effect or if there is a "democracy-promoting" positive effect of oil – then we expect to see more time points clustered in groups 2 and 3. To compute posterior estimates, we use MCMC methods and employ the `R` package `mixAK` by Komárek and Komárková (2013*a*), which uses `Rcpp` to speed up computation

[4] Note that this is also an advantage of the structural breaks approach as well as the dynamic linear modeling framework.

[5] The value of -0.1 is chosen because this represents the average coefficient estimate from Ross (2012), Table 3.1. The value of 0.0 intuitively represents a null effect, and the value of 0.1 is set as a symmetric positive effect. These values are to be set by the researcher according to the clustering hypotheses of interest.

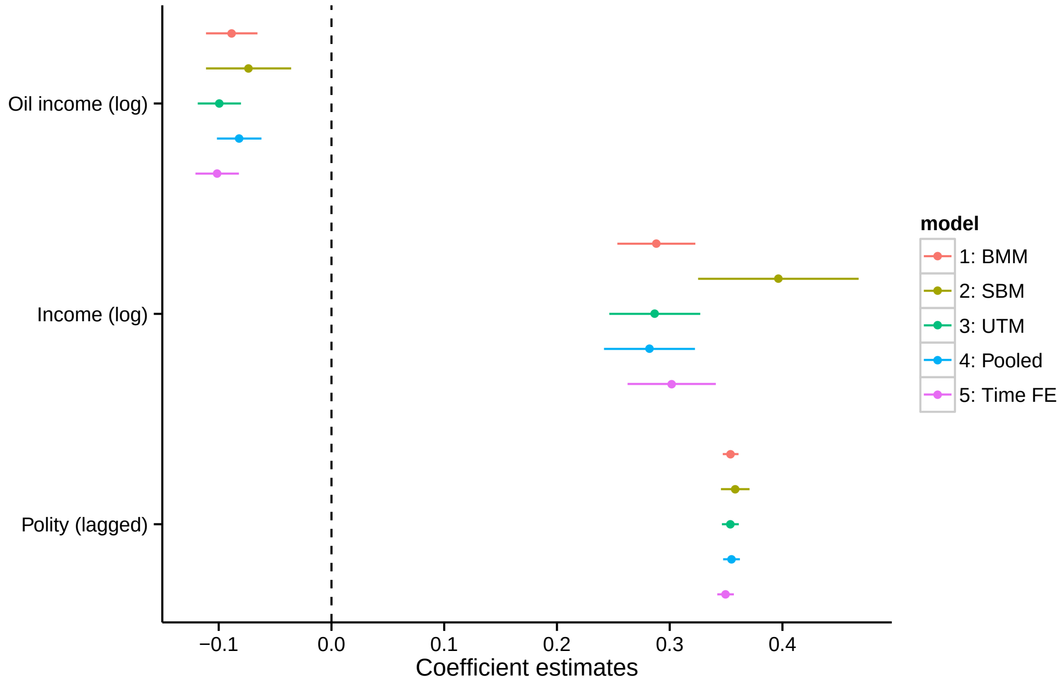

Figure 2: Replication and extension of Ross (2012). Dependent variable is polity score. All models include an AR(5) autocorrelation structure. Abbreviations: Time FE – time fixed-effects model, used in Ross (2012); UTM – unstructured time model; DLM – dynamic linear model; SBM – structural breaks model (with $k = 2$ break points); BMM – Bayesian mixture model (with $k = 3$ clusters). Note that for the SBM, the coefficient estimates are the average fixed-effects estimates across all periods. Similarly, for the BMM, the coefficient estimates for the oil income variable is the expected value of the random effect parameter β_t.

and is designed to augment the familiar `MCMCpack` package (Martin, Quinn and Park 2011). From the MCMC output, we will examine fixed-effect estimates and three quantities of interest: posterior mixture distributions, clustering frequency, and clustering assignments. The appendix provides the `R` code for the mixture model using the `mixAK` package.

To ensure the consistency of fixed-effects parameter estimates across the different model specifications discussed in this paper, we replicate Ross's results in Figure 2. The coefficients' means and uncertainty estimates are consistent across all models for the fixed-effects parameters of national income and 5-year lagged Polity score. Since the oil income variable is a fixed-effect estimate in Ross's original model and the more simple pooled model (not including year fixed-effects, as Ross does), in order to compare the estimates to the other model specifications, we use expectations. For the unstructured time model, we can simply average the $T-1$ time-oil interactions coefficients and use bootstrapped standard errors. Doing this, we get similar coefficient estimates from the unstructured time model (-0.099) compared to the base model (-0.101), though with larger standard errors. For the structural breaks model, we take expectations of the fixed-effects estimates across all break points (that is, the model estimates 3 different estimates if we use $k = 2$ breaks, so we average across the breakpoints in this case). The coefficient estimates are consistent across all model specifications, with one exception: the income variable in the structural breaks model is estimated to be slightly larger than in the other models and with higher uncertainty, but is still positive and statistically significant.

Turning now to the Bayesian mixture model approach, we first assess model convergence. Though not shown here, traceplots and autocorrelation plots of the deviance parameter, mixture means, and fixed effects parameters all indicate proper convergence of the MCMC algorithm. We stress that it is important to examine the diagnostic plots for the mixture parameters thoroughly to ensure convergence of the MCMC procedure. This specifically applies to the mixture weights w and the mixed β_t random effects. Given proper convergence, we then examine fixed effects estimates. As shown in Figure 2, the lagged polity and GDP income coefficient estimates are consistent with results from Ross (2012) and the other models. To obtain a comparable coefficient estimate for oil income, we use the posterior expectation of $\beta_t | z_t, \mu_k, D_k$. This gives an estimate of -0.088 which is consistent with Ross's coefficient estimate of -0.101.

Of course, this estimate is an average of what we have already discerned is a temporally heterogeneous process. The richness of the mixture modeling framework is in the clustering assignment of time points (years) for the parameter of interest, β_t, so that we can observe not only the degree of temporal heterogeneity in effects but also the point estimates conditional on cluster assignment. First, we can visualize the posterior distributions of the cluster means for each of the $K = 3$ groups, which are presented in Figure 3. Though there is some overlap across distributions, overall we see that there are three distinct Gaussian distributions each with their own mean but with similar variance, consistent with the priors discussed above. Substantively, what the clusters represent is a set of three effect-possibilities: a democracy-hindering effect of oil, a null effect, and a democracy-promoting effect.

Second, we can visualize the frequency of clustering assignments into each of the three groups. This step is particularly useful for researchers using the Bayesian mixture model approach to test hypotheses of temporally heterogeneous effects. If there is no temporal heterogeneity of effects, we

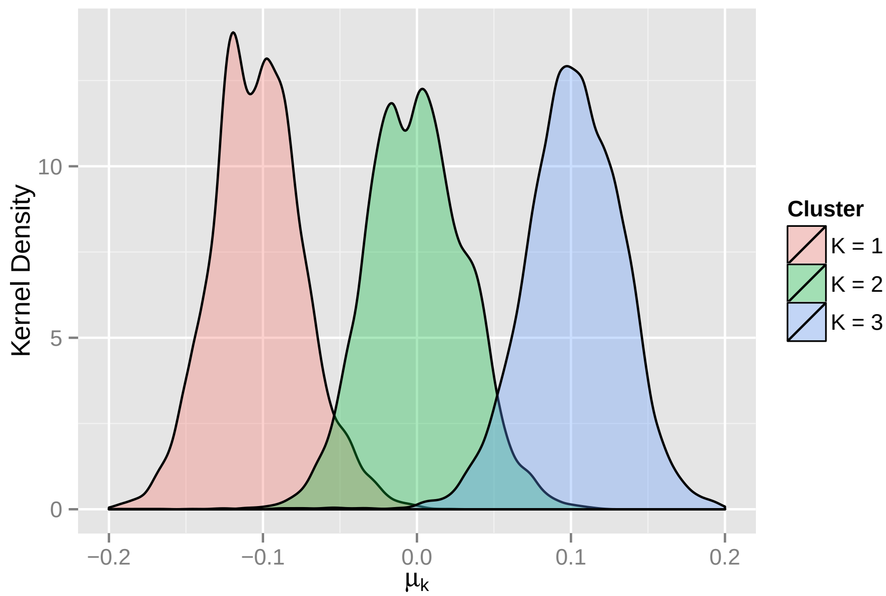

Figure 3: Posterior densities of cluster means based on 200,000 MCMC iterations thinned every 100th draw. Cluster means represent coefficient estimate for effect of oil income on polity scores. Group 1 is assigned a negative estimate (-0.1), Group 2 is assigned a zero (null) estimate (0.0), Group 3 is assigned a positive estimate (0.1). See text for a detailed explanation of cluster means and variances.

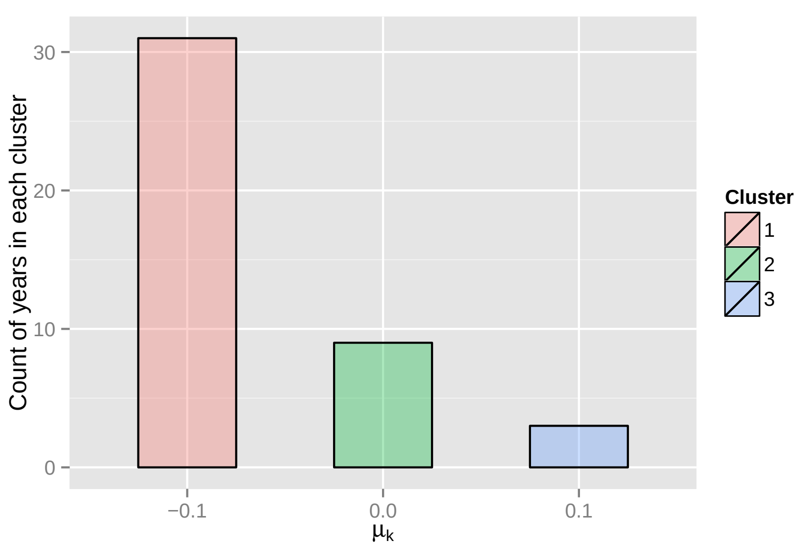

Figure 4: Frequency of posterior cluster assignments of individual time points (years) for each cluster $k \in (1, 2, 3)$. Total number of years in sample is 43. Cluster 1 has negative mean, cluster 2 has zero mean, cluster 3 has positive mean.

will see nearly all years assigned into one cluster. If there is temporal heterogeneity we will see clustering into multiple groups. Recall that if Ross's (2012) claim is supported by the data, most years will be grouped into cluster 1 (with a negative cluster mean); if instead the claims made by Haber and Menaldo (2011) and Dunning (2008) are supported by the data, years will be mostly clustered into cluster 2 (null effect) and cluster 3 (positive effect), respectively. Figure 4 shows the frequency of time points allocated to each of the three clusters. In this illustration, the Bayesian mixture model provides evidence in support of the Ross (2012) conjecture that oil has democracy-hindering effects: of the 43 years in the sample, 31 years for the oil income coefficient are clustered into group 1, with cluster mean equal to -0.1. These findings also provide some support to claims of a null effect of oil on democracy as 9 years are clustered into group 2, though there is little evidence of a democracy-promoting oil effect given that only 3 of the 43 years are clustered into group 3.

Third, we examine the specific allocation of time points into clusters. This is done by assigning years to clusters based on the posterior weighting probabilities w_k. Specifically, time points are placed into the corresponding group k with the highest weighting probability for each time point t (Komárek and Komárková 2013*b*). Cluster assignments can be visualized by applying the cluster means to specific time points, along with estimates of uncertainty, which are assumed constant across clusters (this assumption may be relaxed by allowing $\mathbf{D}_k$ to vary by cluster). Table 1 shows the specific assignment of years to clusters as a text-bar chart. Notice again that very few years were clustered into group 3, suggesting little evidence of a positive effect of oil on democracy.

Results from the posterior clustering probabilities show similar patterns to the coefficient plot presented in Figure 1. A key difference between the Bayesian mixture model and the simpler unstructured time model is the assignment of the early years of the dataset into the category of a negative but small effect. Where the unstructured time model shows estimates of a statistically null effect for 16 years and a negative effect for the remaining 27 years, the mixture model gives results of a null effect for 9 years with a positive effect for the 3 years in the early 1990s (consistent with Dunning's argument for the early 1990s). Compare this to the results of the Bayesian structural breaks model which estimates a negative effect before 1976, a null effect between 1977 and 1989, and a negative effect after 1990.[6] If we think of each period in between breaks as a cluster, the

[6] Specifically, the Bayesian structural breaks model estimates two breaks at $t = 15$ and $t = 28$ with 95% credible

structural breaks model estimates negative effects for 29-31 years and null effects for 12-14 years. Though the structural breaks model does not show estimates for a positive effect, the findings for

k	*Posterior assignment of years in each cluster*
1	61, 63, 64, 65, 68, 69, 70, 71, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 85, 86, 87, 89, 95, 96, 97, 98, 99, 00, 01, 03
2	62, 66, 67, 72, 84, 88, 92, 94, 02
3	90, 91, 93

Table 1: Posterior cluster assignments of individual time points (years) for the effect of oil income (log) on polity scores for each cluster $k \in (1, 2, 3)$. Total number of years in sample is 43. Cluster means are -0.1, 0.0, and 0.1 for the three clusters, with cluster-constant standard error of 0.01.

years in which there is a negative effect and years in which there is a non-negative effect are quite similar to the results from the Bayesian mixture model.

If a more "smooth" clustering of years is preferred when using the Bayesian mixture model, the covariance matrix $\mathbf{D}_k$ can be explicitly modeled with a temporal dependence structure as discussed above. This would reduce the chance of singleton years assigned to clusters, as is the case for 1962, 1972, 1984, 1988, 1992 and 2002. Such an assumption would lead estimates from the Bayesian mixture model to more closely resemble the Bayesian structural breaks model. In this way, the Bayesian mixture model is the most flexible of the approaches considered here and in its most unrestricted form, the framework offers a data-based solution to clustering with little dependency on model selection or on researcher-specified assumptions of the nature and degree of temporal heterogeneity of effects.

5 Discussion

This paper offers a solution to modeling longitudinal data with temporally heterogeneous effects, especially when the estimation of temporal clusters themselves is of scientific interest. Often temporal heterogeneity is either ignored or treated as a nuisance parameter. Common approaches such as time fixed-effects models tend to push researchers toward pooling all their data and focusing on estimating some type of "grand mean" for the entire period under analysis (see Andrew Gelman's work for a review of this debate). Yet, often time-averaged parameters are not of theoretical inter-

intervals of $t \in (15, 16)$ and $t \in (28, 29)$, respectively. Since year 0 is 1961, this corresponds to breaks somewhere between 1976 and 1977 and between 1989 and 1990.

est. In the context of longitudinal data, we may not seek claims of the average effect of X on Y over time, but rather seek to find when this effect is strongest or weakest. Moreover, from a pure modeling perspective, a time-invariant assumption is often misguided – as is the case in time-pooled or time fixed-effects models – by assuming the stationarity of what is a time-varying effect in the data generating process. Ideally, this heterogeneity should be tested rather than assumed. As an illustration, in this paper we discussed an active debate in political economy on the relationship between oil and democracy where temporal heterogeneity is a theoretically contested issue. While there are some methods to address temporal heterogeneity in a variety of contexts – for example, unstructured time models, dynamic linear models, and change point analysis – none of these directly deal with the issue of temporal clustering. As in the case of the oil-democracy debate, sometimes the scientific question at hand calls for not only allowing for temporal heterogeneity in the effects but also for actual clustering. We argue that our method produces a data-driven answer – as opposed to an assumptions-driven answer – and that it can directly test temporally-related hypotheses of theoretical interest.

Substantively, the theoretical debate in the resource curse literature in political science is whether natural resources indeed negatively impact democratic governance. Scholars such as Haber and Menaldo have challenged Ross's earlier findings (Ross 2001) by using different model specifications and so-called "fixes" to unit and time heterogeneity, such as country fixed-effects and year (or decade) fixed-effects dummies and the use of error correction models, duration models, and generalized method-of-moments models. Instead of applying these *ad hoc* model specifications, our goal in this paper is to let the data to speak for themselves. Here, the application of the Bayesian mixture model framework to data from Ross (2012) has provided evidence for the democracy-hindering effects of oil wealth for three-quarters of the years in the sample 1961-2003 with evidence for null or positive effects for the other one-quarter years. These results are consistent with theories proposed by advocates of the resource curse: we should not expect negative effects of resource wealth on democratic governance after collapses in commodity prices – as was the case in the late 1980s and early 1990s. This is primarily due to increases in government debt as state expenditures typically rely on high commodity prices to drive government revenue; when these prices collapse, planned expenditures exceed actual revenues and leaders must either renege on political promises to deliver spending to elites or go into debt in order to meet their spending obligations (Karl 1997). In these scenarios, despite high levels of resource rents relative to other countries, resource revenue does not play the regime-stabilizing role that it does in years prior to a price collapse. This was exactly the case for oil-producing states during the post-1986 years when the market experienced an “oil glut” and the price of crude collapsed from $60 per barrel in 1984-85 to $29 per barrel in 1986 and $26 per barrel in 1993 (Yergin 1991). Indeed, results from the Bayesian mixture model support this argument given that the bulk of years assigned to clusters with non-negative means are the years between 1988 and 1994 when we should not expect strong evidence of a resource curse. In nearly all other years, however, the argument that natural resources hinder democracy is well supported by our findings here.

To conclude, we want to emphasize that our method is a complement rather than a substitute for existing approaches. We hope the methodology we presented here can be useful for future researchers working with longitudinal data, in conjunction with approaches that address temporal heterogeneity such as the Bayesian structural modeling framework. We believe the Bayesian mixture modeling approach should be especially useful when researchers are not only interested in dynamic effects but also in classifying the time points in different clusters, according to the effect of the key variable of scientific interest.

References

Acemoglu, Daron and James Robinson. 2005. *Economic Origins of Dictatorship and Democracy*. New York: Cambridge University Press.

Andersen, Jorgen J. and Michael Ross. 2013. "The Big Oil Change: A Closer Look at the Haber-Menaldo Analysis." *Comparative Political Studies* June:1–29.

Andrews, D.W.K. 1993. "Tests for parameter instability and structural change with unknown change point." *Econometrica* 61:821–856.

Aslaksen, Silje. 2010. "Oil and Democracy: More than a Cross-Country Correlation?" *Journal of Peace Research* 47:421–431.

Beardsley, Kyle and Holger Schmidt. 2012. "Following the Flag or Following the Charter? Examining the Determinants of UN Involvement in International Crises, 1945–20021." *International Studies Quarterly* 56(1):33–49.

Box, G. E. P. and G. C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

Chiozza, Giacomo. 2002. "Is There a Clash of Civilizations? Evidence from Patterns of International Conflict Involvement, 1946-97." *Journal of Peace Research* 39(6):711–734.

Dempster, A. P., N. M. Laird and Donald B. Rubin. 1977. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society B* 39:1–38.

Dunning, Thad. 2008. *Crude Democracy*. Princeton University Press.

Fearon, James and David Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97:75–90.

Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin. 2004. *Bayesian Data Analysis*. Chapman and Hall/CRC.

Haber, Stephen and Victor Menaldo. 2011. "Do Natural Resources Fuel Authoritarianism? A Reappraisal of the Resource Curse." *American Political Science Review* February:1–26.

Karl, Terry Lynn. 1997. *The Paradox of Plenty: Oil Booms and Petro-States*. Studies in International Political Economy Series University of California Press.

Komárek, Arnošt and Lenka Komárková. 2013*a*. "Capabilities of R package mixAK for Clustering Based on Multivariate Continuous and Discrete Longitudinal Data.".

Komárek, Arnošt and Lenka Komárková. 2013*b*. "Clustering for multivariate continuous and discrete longitudinal data." *The Annals of Applied Statistics* 7(1):177–200.

Martin, Andrew D, Kevin M Quinn and Jong Hee Park. 2011. "MCMCpack: Markov Chain Monte Carlo in R." *Journal of Statistical Software* 42(9):1–21.

Martin, Andrew and Kevin Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999." *Political Analysis* 10:134–153.

Musgrave, Paul and Yu-Ming Liou. 2010. "Applying New Methods, New Variables, and New Tests to the Resource Curse Hypothesis." Georgetown Working Paper.

Park, Jong Hee. 2011. "Changepoint analysis of binary and ordinal probit models: An application to bank rate policy under the interwar gold standard." *Political Analysis* 19(2):188–204.

Park, Jong Hee. 2012. "A Unified Method for Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models." *American Journal of Political Science* 56(4):1040–1054.

Ross, Michael L. 2001. "Does Oil Hinder Democracy?" *World Politics* 53(3):pp. 325–361.

Ross, Michael L. 2012. *The Oil Curse: How Petroleum Wealth Shapes the Development of Nations*. Princeton, NJ: Princeton University Press.

Spirling, Arthur. 2007. "Bayesian Approaches for Limited Dependent Variable Change Point Problems." *Political Analysis* 15:387–405.

Sweeney, Kevin and Paul Fritz. 2004. "Jumping on the Bandwagon: An Interest-Based Explanation for Great Power Alliances." *Journal of Politics* 66(2):428–449.

Ulfelder, Jay. 2007. "Natural-Resource Wealth and the Survival of Autocracy." *Comparative Political Studies* 40:995–1018.

West, Michael and Jeffrey Harrison. 1997. *Bayesian Forecasting and Dynamic Models*. New York: Springer.

Western, Bruce and Meredith Kleykamp. 2004. "A Bayesian Change Point Model for Historical Time Series Analysis." *Political Analysis* 12:354–374.

Yergin, Daniel. 1991. *The Prize*. Simon and Schuster.

Appendix

Here we provide sample R code for running the Bayesian mixture model using the mixAK package. Code for diagnostics and extracting posteriors is also included.

```
##########################################################################
#              Bayesian mixture model: mcmc using mixAK                 #
##########################################################################
library(mixAK); library(MCMCpack)

mmod1 <- GLMM_MCMC(y = data.na[,"polity"],
                   dist = "gaussian",
                   id = data.na[,"year"],
                   x = list(data.na[,c("polity_5",
                            "logGDP_cap2000_sup_1")]),
                   z = list(data.na[,"logoil_gas_valuePOP_1"]),
                   random.intercept = FALSE,
                   prior.b = list(Kmax = 3,xi = c(-0.1, 0.0, 0.1),
                                  D=rep(.001,3), zeta=3, gD=0.2, hD=500),
                   nMCMC = c(burn = 20000, keep = 10000, thin = 10,
                             info = 100),
                   parallel = TRUE)



##########################################################################
#             Bayesian mixture model: diagnostics and posteriors         #
##########################################################################
print(mmod1[[1]]$prior.b)
print(mmod1[[1]]$prior.b$"xi")
print(mmod1[[1]]$w_b[1:3,])
apply(mmod1[[1]]$mixture_b,2,mean)
apply(mmod1[[1]]$mu_b,2,mean)
apply(mmod1[[1]]$mu_b,2,sd)
apply(mmod1[[1]]$w_b,2,mean)
```

```
tracePlots(mmod1[[1]], param = "Deviance")
tracePlots(mmod1[[1]], param = "sigma_eps")
tracePlots(mmod1[[1]], param = "alpha")
tracePlots(mmod1[[1]], param = "mu_b")
tracePlots(mmod1[[1]], param = "w_b")
tracePlots(mmod1[[1]], param = "sd_b")

autocorr.plot(mmod1[[1]]$Deviance,lag.max = 20,
              col="blue4",auto.layout = FALSE, lwd = 2)
autocorr.plot(mmod1[[1]]$mu_b,lag.max = 20,
              col="blue4",auto.layout = FALSE, lwd = 2)
autocorr.plot(mmod1[[1]]$alpha,lag.max = 20,
              col="blue4",auto.layout = FALSE, lwd = 2)


##########################################################################
#         Bayesian mixture model: posterior cluster assignments          #
##########################################################################

# Posterior mixture means
mu.k <- mmod1[[1]]$mu_b[,1:3]

# Cluster posterior assignment: frequency
groupMean <- apply(mmod1[[1]]$poster.comp.prob2, 1, which.max)
groupMeans <- apply(mmod1[[1]]$poster.comp.prob2, 2, mean)
pMean <- apply(mmod1[[1]]$poster.comp.prob2, 1, max)
table(groupMean)
groupings <- data.frame(1961:2003,groupMean)
#print(groupings)
```