Game Changers: Detecting Shifts in Overdispersed Count Data*

Matthew Blackwell†

January 4, 2017

Abstract

In this paper, I introduce a Bayesian model for detecting changepoints in a time series of overdispersed count data, such as contributions to candidates over the course of a campaign or counts of terrorist violence. While many extant changepoint models force researchers to choose the number of changepoint *ex ante*, this model incorporates a hierarchical Dirichlet process prior in order to estimate the number of changepoints as well as their location. This allows researchers to discover salient structural breaks and perform inference on the number of such breaks in a given time series. I demonstrate the usefulness of the model with applications to campaign contributions in the 2012 U.S. Republican presidential primary and incidences of global terrorism from 1970 to 2015.

*Thanks to Steve Ansolabehere, Adam Glynn, Gary King, Kevin Quinn, and Maya Sen for comments and suggestions. All remaining errors are my own.

†Assistant Professor, Department of Government, Harvard University, Institute for Quantitative Social Science, 1737 Cambridge Street, Cambridge, MA 02138 (email: mblackwell@gov.harvard.edu, web: mattblackwell.org)

1 Introduction

A common task in the analysis of time-series count data is to estimate any structural breaks in the distribution of the count (Park, 2010; Spirling, 2007). For example, in electoral campaigns, the number of contributions to a given candidate represent a costly form of political participation and, thus, can be seen as a measure of enthusiasm for a particular candidate. Discovering a shift in the distribution of these contributions over time could provide a measure of when a candidacy takes off or falls flat.

To estimate these shifts, I develop a nonparametric Bayesian changepoint model with two important features that make it suitable for handling a wide range of count data such as campaign contributions. First, the model relies on a hierarchical Dirichlet process (HDP) prior to obviate the need to specify the correct number of changepoints *a priori* (Teh et al., 2006). Obviously, for most applications, it would extraordinarily difficult for researchers to know, with certainty, the number of changepoints in the data. For many researchers, in fact, estimating the number of changepoints might be as interesting as estimating their location. The HDP prior allows the model to estimate and perform inference on both the number and the location of changepoints, making it an extremely flexible model for a wide array of applications.

Second, I model the distribution of the counts as negative binomial, which can account for overdispersion in count data. Extant changepoint models for count data in political science (Park, 2010; Spirling, 2007) rely on the Poisson distribution, but many types of counts can have higher variance than a Poisson model would imply. Campaigns often attempt to fundraise through email or at events, both of which lead to clusters of donations at specific times. In counts of terrorist activity, the number of injuries in a particular month might exhibit overdispersion since one attack might produce a large number of injuries and one conflict might produce many attacks. The negative binomial model easily handles these types of overdispersion. These two model features make this a powerful and flexible approach to estimating structural breaks in count data.

2 A Model for Changepoints in Overdispersed Count Data

2.1 Changepoint models

Changepoint models estimate discrete changes in the distribution of time-series data. I focus on a specific class of changepoint models called hidden Markov models (HMMs). Given a time series of observed contribution counts, $y = (y_1, \ldots, y_T)$, an HMM assumes that the count at time t is independent of other time periods conditional on a time-specific state variables, s_t, which follow a Markov process. In the usual finite HMM, there are a finite number of states, $s_t \in \{1, \ldots, K\}$, and each state, $s_t = k$, is associated with a particular set of parameters for the distribution of the outcome, θ_k: $y_t|s_t \sim F(\theta_{s_t})$, where $F(\cdot)$ is a family of distributions. Changepoint models can accommodate many families of data-generating processes, including continuous, binary, and count outcomes (Spirling, 2007; Park, 2010; Park, 2011).

The overdispersion inherent in many types of count data requires a deviation from the commonly used Poisson changepoint models (Chib, 1998; Spirling, 2007; Park, 2010). These models implicitly assume that the mean in any specific regime is equal to the variance, which is unlikely to hold in general and fails in the applications below. As shown by Frühwirth-Schnatter et al. (2009) in the context of mixture modeling, we can handle overdispersion in a count model by augmenting the usual Poisson with a random intercept:

$$y_t|s_t = k, \beta_k, \eta_t \sim \text{Po}(\eta_t \exp(X_t' \beta_k)), \tag{1}$$

where X_t is a $J \times 1$ vector of covariates, β_k are the $J \times 1$ vector coefficients on the covariates from state k, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ is the collection of coefficients across states. If no covariates are included except an intercept term, then each β_k is a scalar. The random effects, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_T)$, allow for the marginal distribution of the data (that is, $p(y_t|\lambda_t)$) to have a separate mean and variance. In fact, if we place a Gamma prior on the random intercept,

$$\eta_t|s_t = k, \rho_k \sim \text{Ga}(\rho_k, \rho_k), \tag{2}$$

then the marginal distribution of the data is negative binomial. Negative binomial models are common in political science for handling count data with overdispersion (King, 1989). Note that the prior in (2) allows for different amounts of overdispersion in different regimes. As ρ_k tends toward infinity, the model converges to a Poisson model.

2.2 Estimating the number of changepoints

A changepoint in an HMM is when the time series transitions from one state to another, so that $s_t \neq s_{t+t}$. Thus, specifying how the model switches in this fashion is important to HMMs in general and changepoint models, specifically. Chib (1998) introduced a Bayesian HMM with a constraint on this transition process so that if $s_t = k$, then s_{t+1} can only stay in state k or transition to a new state, $k + 1$ and there is a known number of regimes, K. In this model, each of these K regimes must be visited so there are exactly $K - 1$ changepoints, which can create misleading estimates if K is misspecified. A common approach to determining the number of changepoints is to estimate many models, each conditional on a number of changepoints, then use a model selection tool to choose the "best" model (Park, 2010; Chib, 1998). These approaches, however, are both technically challenging and place the number of changepoints outside the scope of the model, meaning that there is no way to understand the posterior variability of the number of regimes or changepoints.

To avoid having tp specify the number of changepoints *a priori*, I rely on a Bayesian nonparametric approach called the hierarchical Dirichlet process or HDP (Teh et al., 2006) that allows the model to infer (1) the number of changepoints and (2) their location.[1] The HDP is a generalization of the Dirichlet process prior that creates an *infinite* mixture models as opposed to the *finite* mixture model common in changepoint models.[2] Thus, the Dirichlet process prior places no restrictions on the number of regimes *a priori* (Ferguson, 1973; Escobar and West, 1995). Hierarchical Dirichlet processes allow for different groups of observations to have different mixtures, but to share mixture components (that is, what is being mixed

[1]Other approaches to estimating the number of changepoints exist (Koop and Potter, 2007; Giordani and Kohn, 2008; Geweke and Jiang, 2011; Chong and Ko, 2011). One advantage of the HDP approach is that allows the process to return to previously visited states, which can reduce the number of redundant states and allow for a more meaningful substantive analysis of the regimes.

[2]For other uses of Dirichlet process priors in political science, see Grimmer (2011) and Spirling and Quinn (2010).

over) across groups. In the context of changepoint models and HMMs, the groups are defined by the state, s_t, and the mixtures are the transition probabilities between one state to the next. Using an HDP is crucial in this context, because we want the transition matrices of each state to be mixtures over the same groups of parameters (sometimes called the state space).

The HDP for HMMs (called HDP-HMM) places structure on the transition probabilities from one state to another. Given that the process in state j at time t, we need to determine the probability that the process stays in this state or transitions to a new state, π_j. When there are an infinite number of possible states, this is complicated since π_j is infinite dimensional. Furthermore, each state should have its own set of transition probabilities so that, for instance, the probability of staying in a state is higher than leaving it. Thus, there will be an infinite number of transition probability vectors, π_j. The hierarchical Dirichlet process model handles this by treating these transition probabilities as being drawn from a Dirichlet process prior. One way to represent the HDP-HMM is as a limit of finite hierarchical models:

$$y_t|s_t, \boldsymbol{\beta}, \eta_t \sim \text{Po}(\eta_t \exp(X_t \beta_{s_t})) \tag{3}$$
$$\eta_t|s_t, \boldsymbol{\rho} \sim \text{Ga}(\rho_{s_t}, \rho_{s_t}) \tag{4}$$
$$s_t|s_{t-1} = j, \boldsymbol{\pi}_j \sim \text{Discrete}(\pi_{j1}, \dots, \pi_{jK}) \tag{5}$$
$$\boldsymbol{\pi}_j|\alpha, \boldsymbol{\delta} \sim \text{Dirichlet}(\alpha\delta_1, \dots, \alpha\delta_K), \tag{6}$$
$$\boldsymbol{\delta}|\gamma \sim \text{Dirichlet}(\gamma/K, \dots, \gamma/K). \tag{7}$$

In this model, each current state j has its own vector of transition probabilities to other states, drawn from a Dirichlet distribution, which is itself dependent on a distribution $\boldsymbol{\delta}$ that is also drawn from a Dirichlet. This common distribution allows each of the state-specific distributions to share information and the concentration parameter α controls how similar the $\boldsymbol{\pi}_j$ vectors are to $\boldsymbol{\delta}$. This finite model is equivalent to the HDP-HMM as we let $K \to \infty$. For a richer description of the HDP-HMM and HDPs more generally, see (Teh et al., 2006).

One potential drawback to using such a clustering model for detecting changepoints is that the HDP-HMM will often rapidly switch between different states with the same parameter values (Fox et al., 2011). To avoid these redundant states, I rely on the *sticky* HDP-HMM approach of Fox et al. (2011), which models the transition probabilities with a self-transition bias:

$$\boldsymbol{\pi}_j | \alpha, \boldsymbol{\delta} \sim \text{Dirichlet}(\alpha\delta_1, \ldots, \alpha\delta_j + \kappa, \ldots, \alpha\delta_K) \tag{8}$$

The κ in this derivation is the self-transition bias and will increase the probability of staying in state j, π_{jj}, relative to transitioning to a new state. Thus, the prior means of the transition probabilities are:

$$E[\pi_{jk} | \boldsymbol{\delta}, \alpha, \kappa] = \frac{\alpha\delta_k + \kappa \mathbb{1}\{j = k\}}{\alpha + \kappa}, \tag{9}$$

where $\mathbb{1}\{\cdot\}$ is an indicator function. Note that this model allows the observation process to move back and forth between all states, whereas in most traditional changepoint models, observations can only move "forward" to a new state and cannot "return" to a previous state Chib (1998).

In practice, there is no need to draw parameters for an infinite number of regimes. It is possible to use a *weak limit approximation* with a finite, but large, number of regimes, K (Ishwaran and James, 2001). This will not limit the number of regimes estimated by the model, so long as the upper bound on the number of regimes is large enough to never truncate the distribution in practice. In the applications below, I use such an approximation with $K = 15$, which is sufficient for both applications.

2.3 Quantities of interest

In the Supplemental Materials, I describe a Markov chain Monte Carlo (MCMC) approach to estimating this model. There are several quantities of interest that can be calculated from the MCMC output. First, to measure the location of changepoints, we must find time periods where the latent state switched regimes, which requires some care in the HDP-HMM approach because the regime labels change in each MCMC iteration. As an alternative, I simply calculate the *posterior changepoint probability*:

$$\hat{c}_t = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}(\hat{s}_t^{(m)} \neq \hat{s}_{t-1}^{(m)}), \tag{10}$$

where $s_t^{(m)}$ is the mth MCMC draw of the regime for observation t, and M is the number of MCMC draws. We can calculate this straightforwardly from the MCMC output by finding the proportion of draws where a change occurs at t. More generally, we can calculate the posterior probability that two observation belong to the same regime: $\hat{a}_{jt} = \frac{1}{M}\sum_{m=1}^{M}\mathbb{1}(\hat{s}_j^{(m)} = \hat{s}_t^{(m)})$. A plot of this matrix of values can show where regimes appear to change and when certain regimes are "revisited" in the future. Finally, to avoid the labeling problem for a particular set of regime parameters, I calculate the posterior distribution of the parameters for a given day rather than for a particular regime.

3 Illustrations of the Model

To demonstrate the usefulness of the game-changers model, I apply it to two empirical settings: campaign contributions and terrorist attacks. For both, I use the MCMC algorithm described in the Supplemental Material with 100,000 iterations, thinned by 100, with a burn-in period of 5,000 iterations. In the Supplemental Material, I also present a simulation show how the gamechangers model outperforms a Poisson model with a fixed number of changepoints on overdispersed count data.

3.1 The rise and fall of Herman Cain

The Federal Election Commission (FEC) collects data on contributions of $200 or more to campaigns for federal office made by individuals and groups. The FEC requires campaigns to report several pieces of information, including the date that the campaign received the contribution (Federal Election Commission, 2011). These reports allow researchers to track both the daily number of contributions made to a campaign along with the amount contributed. Unfortunately, extant changepoint models are poorly suited to handle campaign contributions data due to the clustering of both fundraising attempts and contribution processing, both of which lead to overdispersion in the contribution counts. However, these data provide an excellent demonstration of the validity of the model. As an illustration, I consider the candidacy of Herman Cain in the 2012 Republican primary. Cain was one of many candidates vying for the nomination and one of a few to reach the status of frontrunner, quickly losing that status due in part to allegations of sexual

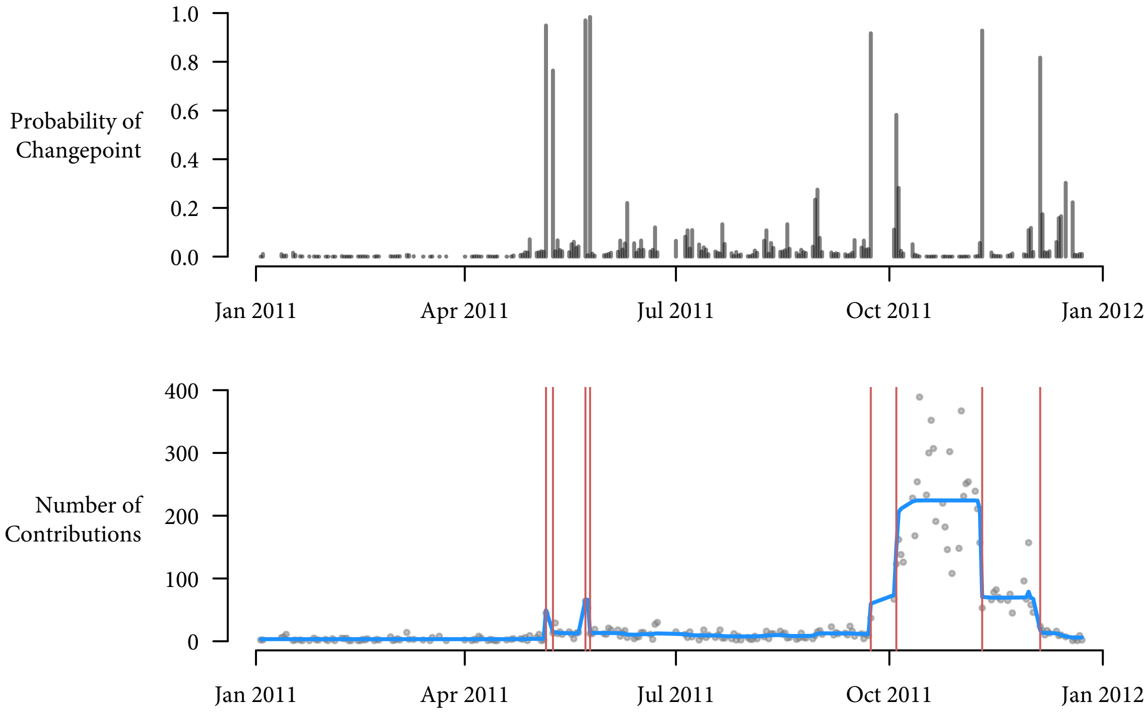

Figure 1: Contributions and changepoints for Herman Cain in the 2012 Republican Primary

misconduct. The quick ups and downs of Cain's campaign provide a good target for the changepoint model.

Figure 1 presents the posterior probability of a changepoint in the top panel.[3] In the bottom panel, I plot the raw number of contributions along with the posterior mean of λ_t, the mean of the negative binomial distribution for each observation in red. The vertical red lines correspond to dates that have greater than 0.5 posterior probability of being a changepoint. Table 1 lists each of these estimated changepoints and its corresponding event in the campaign. The model correctly identify major shifts in the distribution of contributions to Herman Cain that correspond to

[3] I preprocess the data by removing weekends and the days leading up FEC filing dates in order focus on changepoints related to contributor behavior. This is akin to removing stop words when clustering text data.

Changepoint	Pr(Change)	Direction	Event
May 6–9, 2011	0.949/0.764	+/−	Fox News debate (May 5)
May 23–25, 2011	0.97/0.984	+/−	Announces candidacy (May 21)
Sep. 23–Oct. 4, 2011	0.917/0.582	+/+	Wins Florida 5 Straw Poll (Sep. 24)
Nov. 10, 2011	0.928	−	Sexual misconduct allegations (Nov. 7)
Dec. 5, 2011	0.817	−	Suspends campaign (Dec. 3)

Table 1: Estimated Herman Cain changepoints and their substantive explanations.

actual prominent events in his campaign. The model correctly identifies his rise after winning a key straw poll on September 24th, 2011, (Sutton and Holland, 2011) and his fall after sexual misconduct allegations were made public on November 7th (Martin et al., 2011; Henderson, 2011). Note that the model makes no restrictions on the number of changepoints in the data. This is crucial in this example because specifying the number of changepoints *a priori* would be difficult, even if one were to visually inspect the time series.

3.2 Terrorism around the world

Terrorism remains a persistent and malevolent threat in many countries around the world, and how terrorism relates the political world has generated considerable scholarly interest (see Young and Findley, 2011, for a review of this literature). Many of these studies leverage time-series or time-series cross-sectional data on terrorist attacks or the number of injuries due to terrorist attacks. These time series tend to be highly overdispersed, however, since a single attack might induce many clustered injuries or an underlying conflict may lead to "bundles" of attacks in a given country.

To investigate changes in the distribution of terrorist attacks over time, I analyze data from the Global Terrorism Database (START, 2016), which tracks both transnational and domestic terror attacks from 1970 until 2015 (with 1993 missing). I aggregated the number of deaths and number wounded in terrorist attacks to the monthly level to produce a time-series of terrorism-related injuries over 552 months. With this long span of data and quite a few outlier months, allowing the data to choose the number of regimes is vital. Identifying changepoints and common regimes can elucidate some of the root causes of terrorism and point

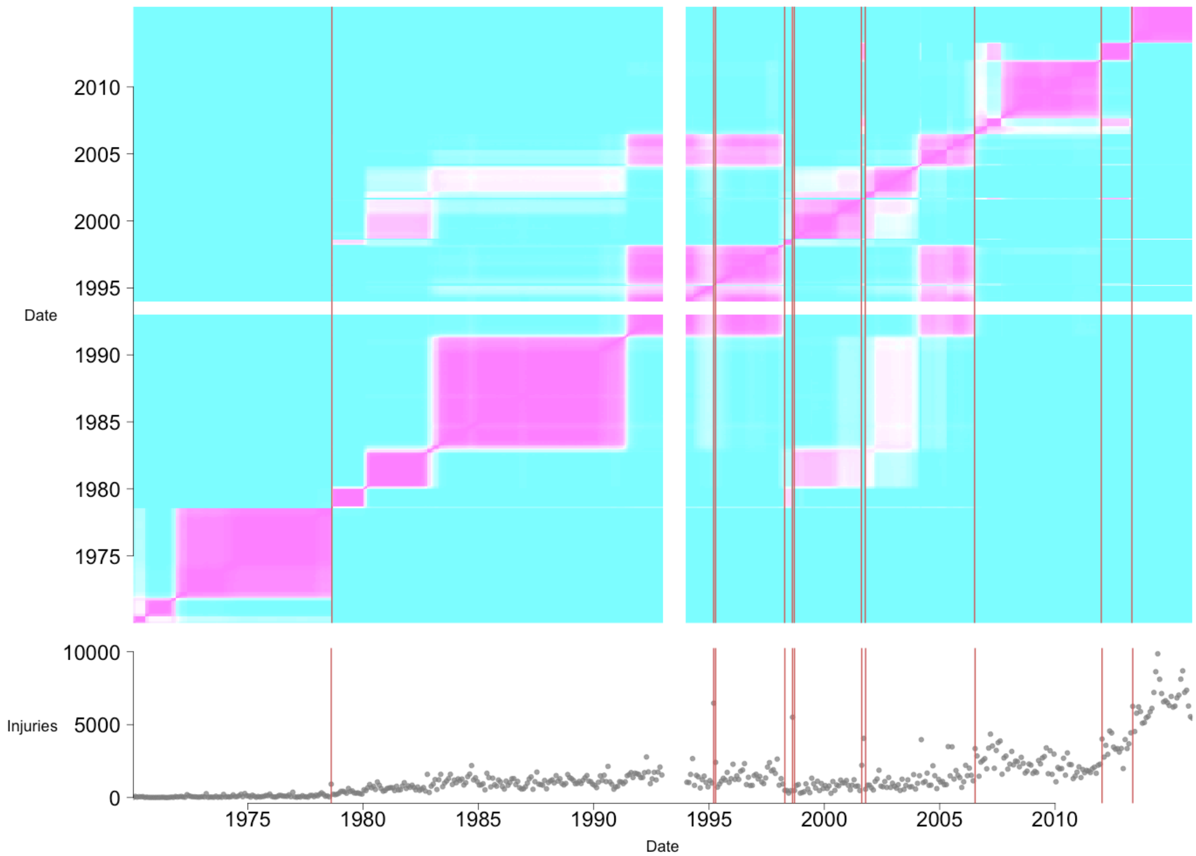

Figure 2: Posterior probability of two months being in the same regime.

researchers to time periods and events worthy of further study.

Figure 2 presents the results of the game-changers model for the terrorism data. The top panel of this figure shows the posterior probability of two time periods being in the same regime, and the bottom panel shows the counts over time. Red vertical lines represent dates with a greater than 0.5 posterior probability of being a changepoint. The clearest message from these results is the relative stability of terrorism in the Cold War era and the relative instability after the USSR collapses in 1991. In the latter era, a few changepoints highlight single months that had an unusually high number of injuries, such as the 9/11 attacks, the August 1998 U.S. embassy bombings, and a combination of the Tokyo sarin gas subway attacks and the Oklahoma City bombing in March and April of 1995, respectively.

There are several changepoints since 9/11, each marking a significant increase in terrorist activity. After June 2006, for instance, the terrorism-induced injury rate in Iraq, Pakistan, India, and Afghanistan increased markedly. Another increase in terrorist attack occurs at the start of 2012, with significantly increased terrorist activity from jihadist groups such as the Taliban (in Afghanistan and Pakistan), Al-Shabaab (East Africa), Al-Qaida in Iraq, and Boko Haram (West Africa). A final regime starts in May of 2013 and continues through the end of the data (late 2015), with increases in activity by all of these groups and the beginning of attacks from the Islamic State of Iraq and the Levant (ISIL). The changepoint that precipitates this final regime coincides with two events. First, there is an escalation of the conflict between Nigeria and Boko Haram. Second, Sunni-Shia violence erupted in May 2013 in reaction to the Iraqi Army raiding an anti-government protest camp in the city Hawija in northern Iraq amid tension surrounding the April parliamentary elections.

Even without guidance on the number or location of structural breaks, the model is able to find politically relevant dates where the distribution of terrorist activity sharply changed. Scholars of terrorism, both quantitative and qualitative, could use this model to better understand when terrorism is likely to take off and when it collapses. Note that extending this model to include continent or region parameters would be straightforward. In this augmented model, changepoints would detect when the overall level of terrorism or the distribution of terrorism across region changes.

4 Conclusion

This paper applies a novel statistical model that estimates the number and timing of changepoints in overdispersed count data. This model, which relies on Bayesian nonparametrics, gives researchers the ability to cluster political time-series into distinct regimes and detect significant shifts in the distribution of the counts. The model uses recent developments in Dirichlet process priors to estimate the number of changepoints rather than specifying the number *a priori*. This is important in many applications where the number of changepoints is unknown or is the target of inference itself. While the model here has been tailored to overdispersed count data, modifying the base (within-regime) model to allow for continuous, binary, and ordered categorical outcome variables is possible.

References

Chib, Siddhartha (1998). "Estimation and comparison of multiple change-point models." *Journal of Econometrics* 86.2, pp. 221–241 (cit. on pp. 3, 4, 6, 15).

Chong, Terence Tai-Leung and Stanley Iat-Meng Ko (2011). "Dirichlet Process Multiple Change-point Model." presented at the Econometric Society Australasian Meeting 2011, Adelaide, Australia (cit. on p. 4).

Escobar, Michael D. and Mike West (1995). "Bayesian Density Estimation and Inference Using Mixtures." *Journal of the American Statistical Association* 90.430, (cit. on p. 4).

Federal Election Commission (2011). *Campaign Guide for Congressional Candidates and Committees* (cit. on p. 7).

Ferguson, Thomas S. (1973). "A Bayesian Analysis of Some Nonparametric Problems." *The Annals of Statistics* 1.2, (cit. on p. 4).

Fox, Emily B et al. (2011). "A sticky HDP-HMM with application to speaker diarization." *The Annals of Applied Statistics* 5.2A, pp. 1020–1056 (cit. on pp. 5, 15, 17).

Frühwirth-Schnatter, Sylvia et al. (2009). "Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data." *Statistics and Computing* 19.4, pp. 479–492 (cit. on pp. 3, 15–17).

Geweke, John and Yu Jiang (2011). "Inference and prediction in a multiple-structural-break model." *Journal of Econometrics* 163.2, pp. 172–185 (cit. on p. 4).

Giordani, Paolo and Robert Kohn (2008). "Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models." *Journal of Business & Economic Statistics* 26.1, pp. 66–77 (cit. on p. 4).

Grimmer, Justin (2011). "An Introduction to Bayesian Inference via Variational Approximations." *Political Analysis* 19.1, pp. 32–47. eprint: `http://pan.oxfordjournals.org/content/19/1/32.full.pdf+html` (cit. on p. 4).

Henderson, Nia-Malika (2011). "Sharon Bialek accuses Herman Cain of sexual harassment as she sought help getting a job." *Washington Post* (cit. on p. 9).

Ishwaran, Hemant and Lancelot F. James (2001). "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association* 96.453, pp. 161–173 (cit. on p. 6).

King, Gary (1989). "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33.3. http://gking.harvard.edu/files/abs/varspecec-abs.shtml, pp. 762–784 (cit. on p. 4).

Koop, Gary and Simon M. Potter (2007). "Estimation and forecasting in models with multiple breaks." *Review of Economic Studies* 74.3, pp. 763–789 (cit. on p. 4).

Martin, Jonathan et al. (2011). "Herman Cain accused by two women of inappropriate behavior." *Politico* (cit. on p. 9).

Neal, Radford M (2003). "Slice sampling." *The Annals of Statistics* 31.3, pp. 705–767 (cit. on p. 17).

Park, Jong Hee (2010). "Structural Change in U.S. Presidents' Use of Force." *American Journal of Political Science* 54.3, pp. 766–782 (cit. on pp. 2–4, 18).

— (2011). "Changepoint Analysis of Binary and Ordinal Probit Models: An Application to Bank Rate Policy Under the Interwar Gold Standard." *Political Analysis* 19.2, pp. 188–204. eprint: `http://pan.oxfordjournals.org/content/19/2/188.full.pdf+html` (cit. on p. 3).

Spirling, Arthur (2007). "Bayesian Approaches for Limited Dependent Variable Change Point Problems." *Political Analysis* 15.4, pp. 387–405. eprint: `http://pan.oxfordjournals.org/content/15/4/387.full.pdf+html` (cit. on pp. 2, 3).

Spirling, Arthur and Kevin Quinn (2010). "Identifying Intraparty Voting Blocs in the U.K. House of Commons." *Journal of the American Statistical Association* 105.490, pp. 447–457. eprint: `http://www.tandfonline.com/doi/pdf/10.1198/jasa.2009.ap07115` (cit. on p. 4).

Study of Terrorism, National Consortium for the and Responses to Terrorism (2016). *Global Terrorism Database [Data file]*. Retrieved from `https://www.start.umd.edu/gtd` (cit. on p. 9).

Sutton, Jane and Steve Holland (2011). "Cain upsets Perry in Florida Republican straw poll." *Reuters* (cit. on p. 9).

Teh, Yee Whye et al. (2006). “Hierarchical Dirichlet Processes.” *Journal of the American Statistical Association* 101.476, pp. 1566–1581 (cit. on pp. 2, 4, 5).
Young, Joseph K and Michael G Findley (2011). “Promise and pitfalls of terrorism research.” *International Studies Review* 13.3, pp. 411–431 (cit. on p. 9).

A Markov Chain Monte Carlo Estimation Strategy

A.1 Priors and hyperparameters

The complete model requires proper priors on all parameters following Frühwirth-Schnatter et al. (2009) and Fox et al. (2011). It is possible to fix the values of the hyperparameters, α, κ, and γ. Following, Fox et al. (2011), I put diffuse priors on these parameters to allow the data to partially determine their value. It is easier to work with transformations of these parameters, $(\alpha + \kappa)$, and $\theta = \kappa/(\alpha + \kappa)$. With these in hand, I use the following independent priors:

$$(\alpha + \kappa) \sim \text{Ga}(1, 0.1); \tag{11}$$

$$\gamma \sim \text{Ga}(1, 0.1); \tag{12}$$

$$\theta \sim \text{Beta}(100, 1); \tag{13}$$

$$\rho_k \propto \rho_k(\rho_k + 10)^{-4}; \tag{14}$$

$$\beta_k \sim \mathcal{N}(0, 25). \tag{15}$$

A.2 Block sampling the latent regimes and HDP parameters

To draw the latent states and the HDP parameters, I use the blocked sampler from Fox et al. (2011). Sampling for the HDP parameters, $\boldsymbol{\delta}$, α, κ, and γ are complicated and require a heavy notational burden, so I refer the interested reader to Appendix D and Appendix E of Fox et al. (2011). To draw the latent states, Fox et al. (2011) rely on a forward-backward procedure similar to the algorithm of Chib (1998). Note that we can write the full conditional posterior of $\mathbf{s}$ as

$$p(s_T|s_{T-1}, \mathbf{y}, \Theta, \boldsymbol{\pi}) \times p(s_{T-1}|s_{T-1}\mathbf{y}, \Theta, \boldsymbol{\pi}) \times \cdots \times p(s_t|s_{t-1}, \mathbf{y}, \Theta, \boldsymbol{\pi}) \times \cdots \times p(s_1|\mathbf{y}, \Theta, \boldsymbol{\pi}), \tag{16}$$

where $\Theta = (\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\eta})$ is the collection of the model parameters. From this derivation, we can see that we can sample s_1 from its full posterior, then sample s_2 conditional on that value of s_2, and so on. To calculate the form of these distribution, however, requires the calculation of a series of "messages" passed from s_t to s_{t-1}.

These messages are defined recursively as:

$$m_{t,t-1}(s_{t-1}) \propto \begin{cases} \sum_{s_t} p(s_t|\pi_{s_{t-1}})p(y_t|\beta_{s_t}, \eta_t)p(\eta_t|\rho_{s_t})m_{t+1,t}(s_t), & t \leq T; \\ 1, & t = T+1; \end{cases} \tag{17}$$

$$\propto p(y_{t:T}|s_{t-1}, \Theta, \boldsymbol{\pi}) \tag{18}$$

With these messages in hand, we can calculate distributions above as:

$$p(s_t|s_{t-1}, \mathbf{y}, \Theta, \boldsymbol{\pi}) \propto p(s_t|\boldsymbol{\pi}_{s_{t-1}})p(y_t|\beta_{s_t}, \eta_t)p(\eta_t|\rho_{s_t})m_{t+1,t}(s_t). \tag{19}$$

With these states in hand, it is straightforward to draw the transition probabilities as a function of the priors and the number of transitions observed in $\mathbf{s}$. That is, I draw

$$\boldsymbol{\pi}_j|\mathbf{s}, \alpha, \kappa, \boldsymbol{\delta} \sim \text{Dirichlet}(\alpha\delta_1 + n_{j1}, \dots, \alpha\delta_j + \kappa + n_{jj}, \dots, \alpha\delta_K + n_{jK})$$

for $j = 1, \dots, K$. Here, n_{jk} is the number of times subsequences in $\mathbf{s}$ with $s_{t-1} = j$ and $s_t = k$.

A.3 Drawing the model parameters

Now that we have draws of the latent states, we need to take draws of the model parameters in each regime (β_k, ρ_k). The non-linear nature of the distributions involved eliminate the possibility of closed-form posterior distributions. This makes the straightforward application of Gibbs sampling impossible. To avoid the inefficiencies of other MCMC approaches, I draw on the auxiliary mixture sampling approach of Frühwirth-Schnatter et al. (2009). This approach augments the data with a set of latent variables τ_{t1} and τ_{t2} which contain all the distributional information about the outcome y and whose distribution can be approximated by a mixture of Normals. With draws of $\tau_t = (\tau_{t1}, \tau_{t2})$ and mixture component indicators $r_t = (r_{t1}, r_{t2})$, we can turn this non-linear problem into a linear Gaussian regression problem. That is, conditional on τ_t, r_t, and η_t, posterior inference on the β_k is simply a Bayesian linear regression. I block sample the negative binomial parameters ρ_k and η_t, using slice sampling to draw ρ_k conditional on $\mathbf{y}$ and $\boldsymbol{\beta}$. With these in hand, ν_t is distributed Gamma with shape $\rho_{s_t} + y_t$ and scale $\rho_{s_t} + \exp(X_t\beta_{s_t})$.

Putting all of these steps together, we have the following draws for a single iteration of the MCMC algorithm:

1. Draw $\mathbf{s}|\mathbf{y}, \Theta, \boldsymbol{\pi}$ as described above.

2. Draw $(\boldsymbol{\rho}, \boldsymbol{\eta})|\mathbf{y}, \boldsymbol{\beta}, \mathbf{s}$:

   a) Draw $\rho_k|\mathbf{y}, \boldsymbol{\beta}$ unconditional on $\boldsymbol{\eta}$ using slice sampling (Neal, 2003).

   b) Draw $\eta_t|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\rho}, \mathbf{s} \sim \text{Gamma}\left(\rho_{s_t} + y_t, \rho_{s_t} + \exp(X_t \beta_{s_t})\right)$, for $t = 1, \dots, T$.

3. Sample $\boldsymbol{\tau}, \mathbf{r}|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\rho}$ using the auxiliary mixture approach Frühwirth-Schnatter et al. (2009).

4. Draw from $\boldsymbol{\beta}|\boldsymbol{\tau}, \mathbf{r}, \boldsymbol{\eta}$ using the auxiliary mixture approach of Frühwirth-Schnatter et al. (2009).

5. Draw $\boldsymbol{\pi}_j|\mathbf{s}, \alpha, \kappa, \boldsymbol{\delta}$ from Dirichlet$(\alpha\delta_1 + n_{j1}, \dots, \alpha\delta_j + \kappa + n_{jj}, \dots, \alpha\delta_K + n_{jK})$, for $j = 1, \dots, K$.

6. Draw $\beta, \alpha + \kappa, \theta$ and γ as in Fox et al. (2011).

A.4 A simulation study

To demonstrate the effectiveness of the gamechangers model, I apply it to a simulated dataset, seen in the bottom panel of Figure 3. This dataset has $T = 200$ observations with four regimes with 50 observations each. I simulated the data in each regime with a simple intercept, so that $\boldsymbol{\beta} = (6, 3, 6, 3)$ and with overdispersion parameters $\boldsymbol{\rho} = (1.5, 0.5, 3, 1.5)$. I ran the above MCMC sampler with an upper bound of 10 states for 10,000 iterations, thinned by 10, with a burin period of 5,000 iterations. Panel (a) of Figure 3 shows the posterior changepoint probabilities and it is clear that there is a high posterior probability of the changepoints occurring around their true values of $t = 51$, $t = 101$, and $t = 151$. Panel (b) of Firgure 3 plots in red the posterior means, $\hat{\lambda}_t$, which track closely to the true values of λ_t in green.

Extant changepoint models in political science also rely on the Poisson distribution, but do not take into account overdispersion. To demonstrate this, I applied

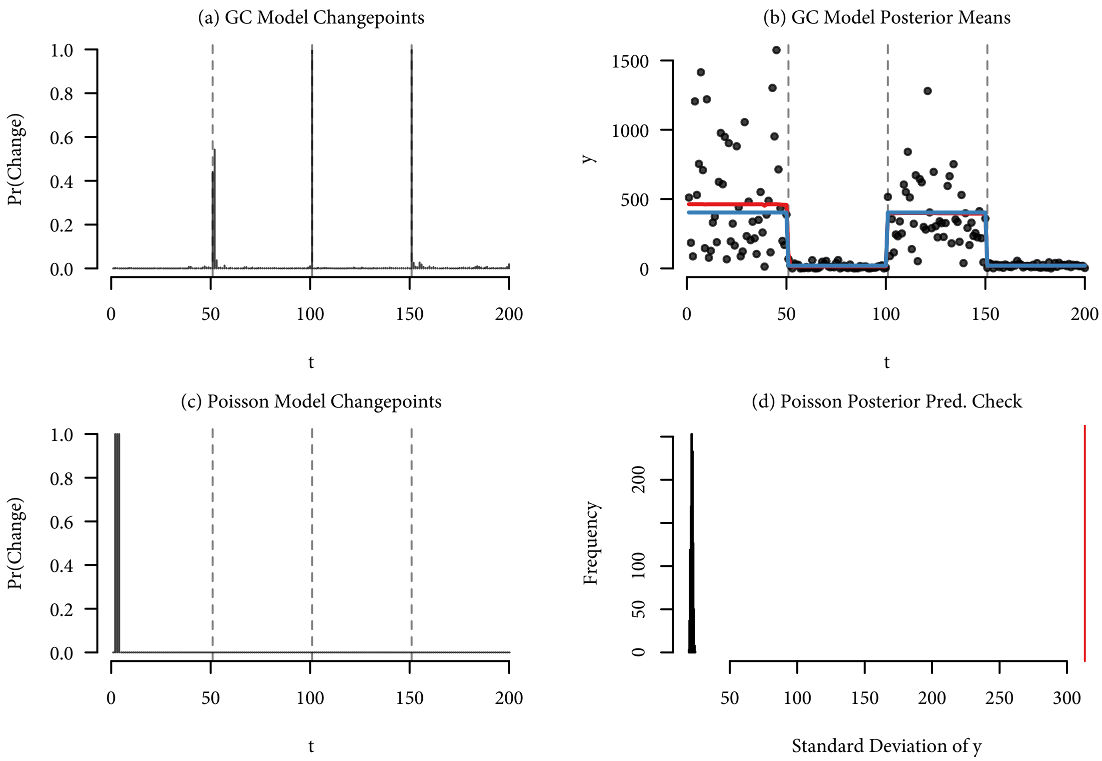

Figure 3: Simulation results for the gamechangers models (a-b) and a naive Poisson model (c-d), with true changepoints at $t = 51, 101$ and 151. Panels (a) and (c) show the posterior probability of a changepoint in each time period for the two models. Panel (b) shows the posterior means by time period (red) and the true means (blue). Panel (d) shows the posterior predictive distribution of the standard deviation of y under the Poisson model and the observed value (red line).

the Poisson changepoint model of Park (2010) to the same set of simulated data. For this model, we must specify the number of changepoints, so to give an advantage, I correctly specify the number of changepoints. Even with this advantage, the Poisson model is unable to recover the true locations of the changepoints. Panel (c) of Figure 3 shows that none of the estimated changepoints come close to the true changepoints. Panel (d) of the same figure shows why the Poisson model fails to find these changepoints. This panel plots a posterior predictive check for overdispersion, which the model clearly fails. This plot shows a histogram of the standard deviations of data predicted by the posterior distribution of the parameters, along with the actual standard deviation of the data in red. Obviously the true standard deviation is considerably higher than what is predicted by the model. This is a clear indication that Poisson changepoint models have difficulty in situations where count data is overdispersed, such as with campaign contributions data.