# Game-changers: Detecting shifts in the flow of campaign contributions*

Matthew Blackwell†

October 23, 2012

**Abstract**

In this paper, I introduce a Bayesian model for detecting changepoints in a time-series of contributions to candidates over the course of a campaign. This *game-changers model* is ideal for campaign contributions data because it allows for overdispersion, a key feature of contributions data. Furthermore, while many extant changepoint models force researchers to choose the number of changepoint *ex ante*, the game-changers model incorporates a Dirichlet process prior in order to estimate the number of changepoints along with their location. I demonstrate the usefulness of the model in data from the 2012 Republican primary and the 2008 U.S. Senate elections.

## §1 Introduction

Electoral campaigns are the central events in the political life of democracies. And, increasingly, campaigns are as much about garnering money as they are about garnering votes. Indeed, candidates view fundraising as a vital and time-consuming part of what they do. For citizens, campaign contributions represent a costly form of political participation. This participation certainly depends on features of the individual (Verba, Schlozman, and Brady 1995), but it also ebbs and flows throughout the election season in response to news coverage, campaign events, and changes in candidate strategy (Mutz 1995). While there is some evidence in political science that momentum matters (Bartels 1985), there have been few studies that attempt to statistically pinpoint *when* campaigns take off or

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

†Department of Government, Harvard University. 1737 Cambridge St K351, Cambridge, MA 02138. web: http://www.mattblackwell.org email: mblackwell@iq.harvard.edu

fall flat. This paper seeks to do just that: find points in time when contributions to a candidate change dramatically.

To estimate these shifts, I propose a novel Bayesian changepoint model, called the *game-changers* model, tailored to handle campaign contribution data. The number of contributors to a campaign on a given day is highly overdispersed due to the clustered processing of contributions and the intermittent nature of political attention. Extant changepoint models for count data, such as those used in political science (Park 2010; Spirling 2007), use the Poisson distribution, which is problematic here because it places inappropriate restrictions on the variance of the data. As shown below, this can lead to incorrect inferences on the location of changepoints. The game-changers model uses random effects to handle overdispersion, an approach that is equivalent to assuming a negative binomial likelihood, which is common in political science (King 1989).

Most changepoint models require researchers to specify the number of changepoints in advance, but it is hardly clear what the "correct" number of game-changers is for any given campaign, let alone a series of campaigns. To alleviate this problem, the game-changers model takes a Bayesian nonparametric approach and estimates the number of changepoints along with their location. This approach is an extension of the Chib (1998) method for estimating changepoints and incorporates a Dirichlet process prior for the clustering of the contributions into regimes. This provides a computationally efficient and conceptually straightforward method for allowing the model to include the number of regimes and, thus, the number of changepoints. While this model is tailored to estimating changepoints for campaign contributions, it can be applied to any time series of overdispersed counts, of which there are many. More generally, the Dirichlet process prior approach to estimating the number of changepoints generalizes beyond this model or even count data.

The paper proceeds as follows. Section 2 describes the campaign contributions data and the various factors that lead to overdispersion. Section 3 describes the game-changers model and the computational approach to fitting the model. Section 4 describes four vignettes that show how the model works in simulated and real data. Section 5 concludes.

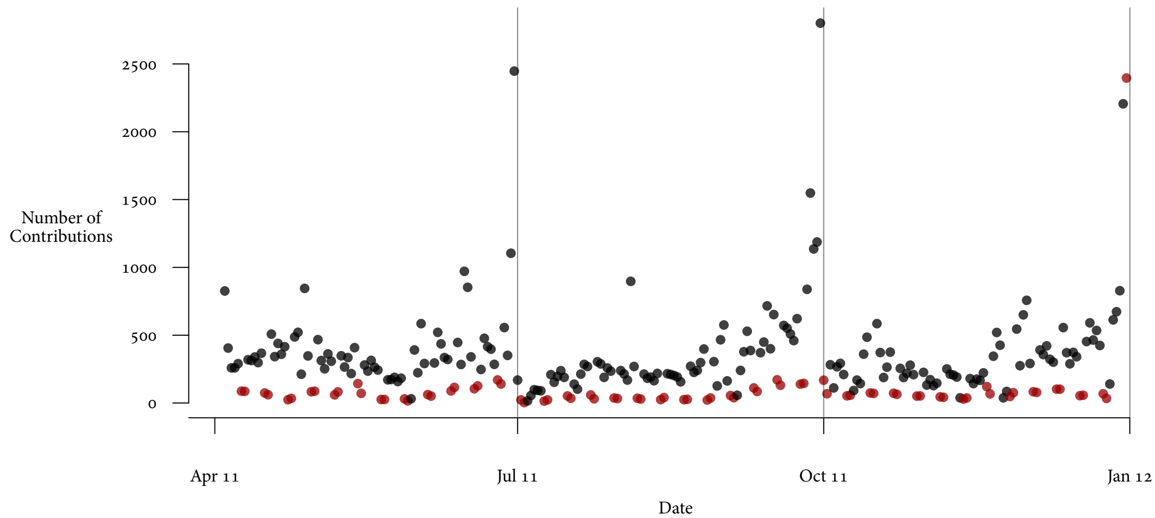


*Figure 1:* Daily number of individual contributions to Barack Obama in 2011. Black dots are weekdays and red dots are weekends. Vertical lines are the filing deadlines.

## §2 The dynamics of campaign contributions

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 a fair amount of information, including the date that the campaign received the contribution (Federal Election Commission 2011). These reports allow us to track both the daily number of contributions made to a campaign along with the amount contributed.

Campaign contributions have a few unique features that make it difficult to apply commonly-used changepoint models. For non-electronic contributions, the date the campaign reports "receiving" a contribution might differ from the date that the date that the donor made the contribution. There are many reasons for this, but two stand out. First, if contributions travel by mail, they might take some time to reach the campaign. Second, and more important for this study, is the potential for delays in campaign processing of contributions. Campaign staff are limited in the amount of time they can process incoming contributions–even if a contribution arrives on a given day, it might not be processed until later. One indication of this is given in Figure 1, which shows the number of contributions received by Barack Obama in 2011, with weekends plotted in red. Campaigns are much more likely to receive contributions on a weekday compared to the weekend. This pattern results from the fact that campaign staff largely work a traditional work week and so contributions that arrive during the weekend are processed after staffers return to work on Monday.

Candidates also have strategic reasons for processing contributions at different rates, one due to signaling and one due to contribution limits (Christenson and Smidt 2011). First, the FEC requires that candidates report their contributions to the FEC at various points throughout the campaign. These reports are important as they publicly disclose the candidate's ability to raise funds. Candidates want to signal that they are a high-quality candidate and one way to do this is to have a large number of contributors. Thus, campaign staff work to process any incoming contributions before these filing deadlines so as to maximize the reported contributions. Figure 1 shows the filing deadlines for 2011 as vertical lines. Clearly, there is a marked increase in the number of contributions received around the filing deadlines. A second reason for pre-deadline increases is that there are different contribution limits for before and after the primary election. A candidate would want to process any pre-primary contributions before the relevant filing deadline so that those pre-primary donors can legally contribute to the campaign again during the run up to the general election. Filing deadlines and weekends are two features campaign contributions data that contribute to the overdispersion of their distribution.

The above reasons for overdispersion could be measured and accounted for in a Poisson regression model, which would alleviate some of the problem. There are other features of campaign contributions that can lead to overdispersion as well, some of which are hard to measure. For instance, campaigns receive many contributions as part of campaign fundraising events—dinners, speaking engagements, and so on. These events add to the clustering of the contributions because they group contributors together in time. These events are more problematic than weekends and filing deadlines because it is very difficult to collect data on the timing of campaign fundraisers. Thus, it is important that we build a model that can handle these unmeasured forms of overdispersion inherent in contributions data.

## §3 A model for changepoints in campaign contributions

### 3.1 Changepoint models

Changepoint models estimate discrete changes in the distribution of time-series data. Given a time-series of observed contribution counts, $Y = (y_1, \ldots, y_T)$, a changepoint model assumes that the distribution of $y_t$ is distributed according to a parameter $\gamma_t$, which takes on at most $M + 1$ distinct values, depending on $t$, where $M$ is the number of changepoints. Thus, $M + 1$ is the number of distinct parameter regimes in the data. Let $\mathbf{c} = (c_1, \ldots, c_M)$ be the vector of changepoints and $\theta = (\theta_1, \ldots, \theta_{M+1})$ be the vector of parameters associated with each regime. With these, we can define the parameters at each point in time as

$$\gamma_t = \theta_m \qquad \text{i.f.f.} \qquad c_{m-1} < t \leq c_m, \tag{1}$$

where we define $c_0 = 0$ and $c_{M+1} = T$. Thus, each observation takes on the parameters of its regime.

One way to conceptualize this model is to imagine the time-series as residing in one regime for a given amount of time before jumping to another regime at a changepoint. Chib (1998) shows that we can think of this regime-switching structure as a discrete-time, discrete-state Markov process with a constrained transition matrix. Let $S = (s_1, \ldots, s_T)$ be a vector of regime indicator, so that if $s_t = m$, then at time $t$ the time-series is in regime $m$ and that $c_{m-1} < t \leq c_m$. Given the nature of the model, we only have to specify the probability of transitioning to the next regime: $\Pr(s_{t+1} = j+1 | s_t = j) = p_{j,j+1}$. We can model $S$ in place of the changepoints since the $k$th changepoint happens at $c_k$ if and only if $s_{c_k} = k$ and $s_{c_k+1} = k + 1$. The regime indicators are useful in Bayesian changepoint models, where we can augment a model with these latent variables to ease computation (Chib and Greenberg 1996).

This model of Chib (1998) forces the time-series to reside in each of the $M+1$ regimes without skipping a regime or returning to a regime. Note though, that if we are interested in estimating the changepoints, $\mathbf{c}$, then recurrent regimes are straightforward since the model will recover the relevant changepoints and treat these recurrence as distinct regimes. More troubling is the lack of regime skipping, which means that each of the $M+1$ regimes is visited. This can be problematic if the true number of structural breaks is less than the number of changepoints in the model. I address this issue below.

### 3.2 *Tailoring changepoints for campaign contributions data*

Up to this point, I have left the distribution of $y_t$ unspecified since changepoint models can accommodate many different data-generating processes, including continuous, binary, and count outcomes. See Park (2010, 2011) and Spirling (2007) for different applications of changepoint models in political science. Unfortunately, the extant changepoint models are poorly suited to handle campaign contributions data due to the features discussed above.

The overdispersion inherent in campaign contributions data requires a deviation from the Poisson changepoint models of Chib (1998), Park (2010), and Spirling (2007). These models assume that

$$y_t|\lambda_t, s_t = k \sim \mathrm{Po}(\lambda_t), \qquad \lambda_t = \exp(X_t\beta_k), \tag{2}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{M+1})$ are the Poisson regression coefficients from each regime. Given the nature of the Poisson distribution, these models implicitly assume that the mean in any specific regime is equal to the variance. This assumption is unlikely to hold in general and fails miserably in campaign contributions data (see Section 4.1 for a demonstration of this).

As shown by Frühwirth-Schnatter et al. (2009) in the context of mixture modeling, we can handle overdispersion in a count model by augmenting model 3 with a random intercept:

$$y_t|\lambda_t, \eta_t, s_t = k \sim \text{Po}(\eta_t \lambda_t), \qquad \lambda_t = \exp(X_t \beta_k). \tag{3}$$

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,

$$\eta_t|s_t = k, \rho_k \sim \text{Ga}(\rho_k, \rho_k), \tag{4}$$

then the marginal distribution of the data is negative binomial. Note that the prior in (4) allows for different amounts of overdispersion in different regimes. As $\rho_k$ tends toward infinity, the model converges to a Poisson model. For a given finite value of $\rho_k$, the marginal distribution of the data has the following form:

$$p(y_t|\lambda_t, \rho_k, s_t = k) = \binom{\rho_k + y_t - 1}{\rho_k - 1} \left( \frac{\rho_k}{\rho_k + \lambda_t} \right)^{\rho_k} \left( \frac{\lambda_t}{\rho_k + \lambda_t} \right)^{y_t}, \tag{5}$$

which is a negative binomial with trial size $\rho_k$ and probability of success $\rho_k/(\rho_k + \lambda_t)$. Negative binomial models are common in political science for handling count data with overdispersion (King 1989).

### 3.3 *Estimating the number of changepoints*

In order to estimate the location of the changepoints, most existing changepoint models require we know the *number* of changepoints that exist in the data. Obviously, for almost any campaign, it would extraordinarily difficult for researchers to know, with certainty, the number of changepoints in the data. For most researchers, in fact, estimating the number of changepoints might be as interesting as estimating their location. A common approach in changepoint models 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).

Changepoint models, though, are a special type of finite-mixture model and these types of models fail to meet the regularity conditions of the traditional, likelihood-based non-nested model comparison tests. Therefore, a common way to compare models is to use Bayesian model selection via the calculation of the marginal likelihood of the model. Park (2011) provides an example of how this approach works for binary and ordinal-probit changepoint models. Chib (1995) provides a straightforward approach to calculating marginal likelihoods when using MCMC based on the Gibbs sampler. This approach is not applicable with the above negative binomial model, however, because it requires a Metropolis-Hastings step to draw the $\rho_k$. Alternative approaches to Bayesian model comparison are computationally difficult and pose problems with highly unlikely models (Park 2011, p. 192). Koop and Potter (2009) identify another major problem with fixed in-sample changepoint approaches: common Bayesian priors, such as those used in Chib (1998), lead to undesirable behavior at the end of the sample.

An alternative to model selection is to estimate the number of changepoints as part of the model itself. A number of methods have been proposed to leave the number of changepoints unrestricted, but many of these approaches are based on a conditionally linear model and not appropriate for the above non-linear model.[1] Instead, this paper preserves the simplicity and computational efficiency of the method proposed by Chib (1998) but allows it to choose the number of changepoints as part of the model.

The approach of Chib (1998) assumes that there are $M+1$ regimes and that each of these regimes is visited by the time-series. The model imposes this restriction by assuming that $s_1 = 1$ and that $s_T = M+1$. This paper instead places no restriction on the value $s_T$, so that the model can estimate

1. Giordani and Kohn (2008) provide a method of estimating the number of changepoints that work for conditionally linear, Gaussian models. Geweke and Jiang (2011) and Chong and Ko (2011) provide alternative MCMC implementations of process priors in changepoint models. Koop and Potter (2007) amends Chib's method to allow for the estimation of the number of regimes, but this approach requires many more calculations than the present approach.

fewer than $M+1$ regimes in the observed sample. This shifts $M$ from being the assumed number of changepoints to the maximum number of changepoints allowed by the model. This approach will recover the posterior distribution on the number of changepoints, as long as we set $M$ high enough not to truncate the posterior. Note that we can only observe $T$ possible regimes in the data—one for each observation.

We can represent this approach as using a specific version of the Dirichlet process prior, a popular tool in Bayesian nonparametrics (Neal 2000). The Dirichlet process prior creates an *infinite* mixture model as opposed to the *finite* mixture models that are typically used by changepoint models.[2] In general, models with Dirichlet process priors group observations together into a countably infinite set of groups (Ferguson 1973; Escobar and West 1995). We can show the central intuition of the Dirichlet process prior as by taking the limit of finite mixture models. Suppose we have a mixture model with the same models as above and $K$ components:

$$
\begin{aligned}
y_t | s_t, \boldsymbol{\beta}, \boldsymbol{\rho}, \eta_t &\sim \text{Po}(\eta_t \exp(X_t \beta_{s_t})) &(6)\\
s_t | \mathbf{p} &\sim \text{Discrete}(p_1, \ldots, p_K) &(7)\\
(\beta_k, \rho_k) &\sim G_0 &(8)\\
\mathbf{p} &\sim \text{Dirichlet}(b/K, \ldots, b/K). &(9)
\end{aligned}
$$

Here, $G_0$ is the "base" distribution of the regime parameters. Neal (2000) shows that we can marginalize over the distribution of $\mathbf{p}$ and, as $K \to \infty$, we find that:

$$
\begin{aligned}
p(s_t = k | s_1, \ldots, s_{t-1}) &\to \frac{n_{t,k}}{t-1+b} &(10)\\
p(s_t \neq s_j \text{ for all } j < t | s_1, \ldots, s_{t-1}) &\to \frac{b}{t-1+b} &(11)
\end{aligned}
$$

Here $n_{t,k}$ is the number of observations up to time $t$ are in component $k$. Thus, each observation

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

is allocated to a component with a probability that is proportional to the number of previous units already allocated to that component. This property of the Dirichlet process prior is called the "rich get richer" property and is a fundamental assumption of the prior. Different Bayesian nonparametric priors have different assumptions embedded into their design and these different assumptions can lead to different clusterings. With this prior in hand, Neal (2000) provides a host of MCMC algorithms to estimate the posterior distribution of both the clusters and the cluster parameters.

A changepoint model is a special case of a clustering, where we refer to the clusters as regimes and restrict how the observations move from regime to regime. Namely, we stipulate that an observation at time $t$ must either be in the same regime as observation $t-1$ or it can form a new regime. Observations cannot "return" to a previous regime. Thus, the mixing probabilities $\mathbf{p}$ do not follow the symmetric Dirichlet distribution of (9). For $s_{t+1}$, all $p_k$ are 0 with the exception of $p_{s_t}$ and $p_{s_t+1} = 1 - p_{s_t}$. These are the probability of remaining in the same regime as $t$ and the probability of moving to a new regime. Since there are only two possibilities, our prior over these values becomes a Beta distribution with parameters $a$ and $b$. This setup implies a Dirichlet process prior with the following transition probabilities as $K \to \infty$:

$$p(s_t = k | s_{t-1} = k, s_1, \ldots, s_{t-2}) \to \frac{n_{t,k} + a}{n_{t,k} + a + b} \tag{12}$$

$$p(s_t = k+1 | s_{t-1} = k, s_1, \ldots, s_{t-2}) \to \frac{b}{n_{t,k} + a + b}. \tag{13}$$

Note that these transition probabilities are no longer Markovian, as they are in the original Chib (1998). This only requires a modest adjustment to the algorithm to draw the $s_t$.

In practice, there is no need to draw parameters for an infinite number of regimes. Instead of sampling from the infinite mixture model, I take an alternative approach that uses a truncated approximating distribution with a finite, but large, number of regimes (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 empirical examples below, I use an upper bound of 20 changepoints and there is never more than 11 changepoints estimated in any iteration of the MCMC algorithm.

### 3.4 *Priors and hyperparameters*

The complete model requires proper priors on all parameters and I use the following:

$$\rho_k \propto \rho_k^{e-1}(\rho_k + d)^{e+f}; \tag{14}$$

$$\beta_k \sim \mathcal{N}(0, B_0); \tag{15}$$

$$p_{k,k+1} \sim \text{Beta}(a, b). \tag{16}$$

The prior for each regime parameters are *a priori* independent. In order for the posterior to exist, the priors must be proper, which means that $e > 1$ for the prior on $\rho_k$. For all of the models below, I use $e = f = 2$ and $d = 10$, which follows Frühwirth-Schnatter et al. (2009), and $B_0 = 100$.

The priors on $p_{k,k+1}$ imply a prior on the length of each regime and, therefore, a prior on the number of regimes that are visited in the sample. Namely, $p_{k,k+1}$ is the probability of a one-period regime, which we can build up to infer an expected *a priori* regime length. In the applications below, I use $a = 20$ and $b = 0.1$, which implies an expected regime length of 200 days and around 1.5 regimes observed in a typical election season. These priors are intentionally designed to allow for long regimes and potentially no changepoints at the expense of finding shorter regimes. When we assume shorter regimes *a priori*, we end up identifying clusters of one- or two-day outliers in addition to the more clearly "game-changing" changepoints. In any case, the estimated changepoints do not vary too much as we change the value of the hyperparameters $a$ and $b$.

### 3.5 *A Markov chain Monte Carlo estimation strategy*

Given the above model, we can write the posterior as follows:

$$
\begin{aligned}
p(\mathbf{s}, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\eta} | \mathbf{y}, \mathbf{X}) \propto\ & p(y_1 | \beta_1, \eta_1, X_t) p(\eta_1 | \rho_1) \times \\
& \prod_{t=2}^{T} \left[ \sum_{m=1}^{M+1} p(y_t | \beta_m, \eta_t, X_t) p(\eta_t | \rho_m) p(s_t = m | \beta_m, \rho_m) \right] \times \\
& \prod_{i=1}^{M+1} p(\beta_m | B_0) p(\rho_m | d, e, f) p(p_{i,i+1} | a, b)
\end{aligned} \tag{17}
$$

To sample from this, I take a Markov chain Monte Carlo approach using Gibbs sampler which samples from the full conditional posterior of each parameter. Below, I discuss the non-standard steps in detail.

### 3.5.1 Drawing the latent regimes

To draw the latent states, I use a modified version of the Chib (1998) algorithm. Chib points out that we can write the full conditional posterior of $\mathbf{s}$ as

$$
p(s_T | \mathbf{y}, \Theta, P) \times p(s_{T-1} | \mathbf{y}, s_T, \Theta, P) \times \cdots \times p(s_t | \mathbf{y}, \mathbf{s}_{t+1:T}, \Theta, P) \times \cdots \times p(s_1 | \mathbf{y}, \mathbf{s}, \Theta, P), \tag{18}
$$

where $\Theta = (\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\eta})$ is the collection of the model parameters, $P = (p_{1,2}, \ldots, p_{M,M+1})$ is the collection of transition probabilities, and $\mathbf{s}_{t+1:T} = (s_{t+1}, \ldots, s_T)$. Crucially, note that Chib (1998) drops the term for $s_T$ because Chib assumes the last observation is in the last regime, $M+1$, with probability one. In this specification, we allow $s_T$ to take any value between 1 and $M+1$, with a probability determined by the data. With this in hand, we can derive each of these distribution and then sample from each, in turn:

- $s_T$ from $p(s_T | \mathbf{y}, \Theta, P)$,
- $s_{T-1}$ from $p(s_{T-1} | \mathbf{y}, s_T, \Theta, P)$,
- $\vdots$

- $s_2$ from $p(s_2|\mathbf{y}, \mathbf{s}_{3:T}, \Theta, P)$.

The regime of the first period is always $s_1 = 1$. Thus, to sample from this, it is sufficient to sample from $p(s_t|\mathbf{y}, \mathbf{s}_{t+1:T}, \Theta, P)$, which is given by Chib (1998).

### 3.5.2 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 $(\beta_k, \rho_k)$. 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 $\tau_{t1}$ and $\tau_{t2}$ 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 $\tau_t$, $r_t$, and $\eta_t$, posterior inference on the $\beta_k$ is simply a Bayesian linear regression. Frühwirth-Schnatter et al. (2009) also shows how to include draws for the negative binomial parameters $\rho_k$ and $\eta_t$ in a Gibbs sampler.

### 3.5.3 MCMC algorithm

Thus, I proceed to draw from the posterior using the following Gibbs sampling approach:

1. Draw $\mathbf{s}|\mathbf{y}, \Theta, P$ as in Section 3.5.1.
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 a Metropolis-Hastings step.
   (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 $p_{ii}|\mathbf{s}, a, b$ from $\text{Beta}(a + n_{ii}, b + 1)$, for $i = 1, \ldots, M + 1$.

# §4 Vignettes

### 4.1 *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 2. 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 20 changepoints for 10,000 iterations, thinned by 10, with a burin period of 5,000 iterations.

The nonparametric nature of the sampler makes visualizing the posterior more complicated than in more traditional approaches to changepoint problems. Namely, since the number of regimes can change from iteration to iteration, it makes little sense to look at the probability of a given observation residing in a specific regime—the nature of the regimes themselves are changing. An alternative approach is to simply calculate the *posterior changepoint probability*, which is simply

$$\hat{c}_t = \frac{1}{G} \sum_{g=1}^{G} \sum_{j=1}^{M} \mathbb{I}(\hat{s}_t^{(g)} = j + 1, \hat{s}_{t-1}^{(g)} = j), \tag{19}$$

where $\mathbb{I}()$ is an indicator function and $s_t^{(g)}$ is the $g$th draw of the regime for observation $t$. We can calculate this straightforwardly from the MCMC output by finding the proportion of draw where a change occurs at $t$. The top panel of Figure 2 shows these values for the simulated data. It is clear that there is a high posterior probability of the changepoints occurring around their true values of $t = 51$, $t = 101$, and $t = 151$.

Making inferences about the regime parameters is also difficult due to the changing number of

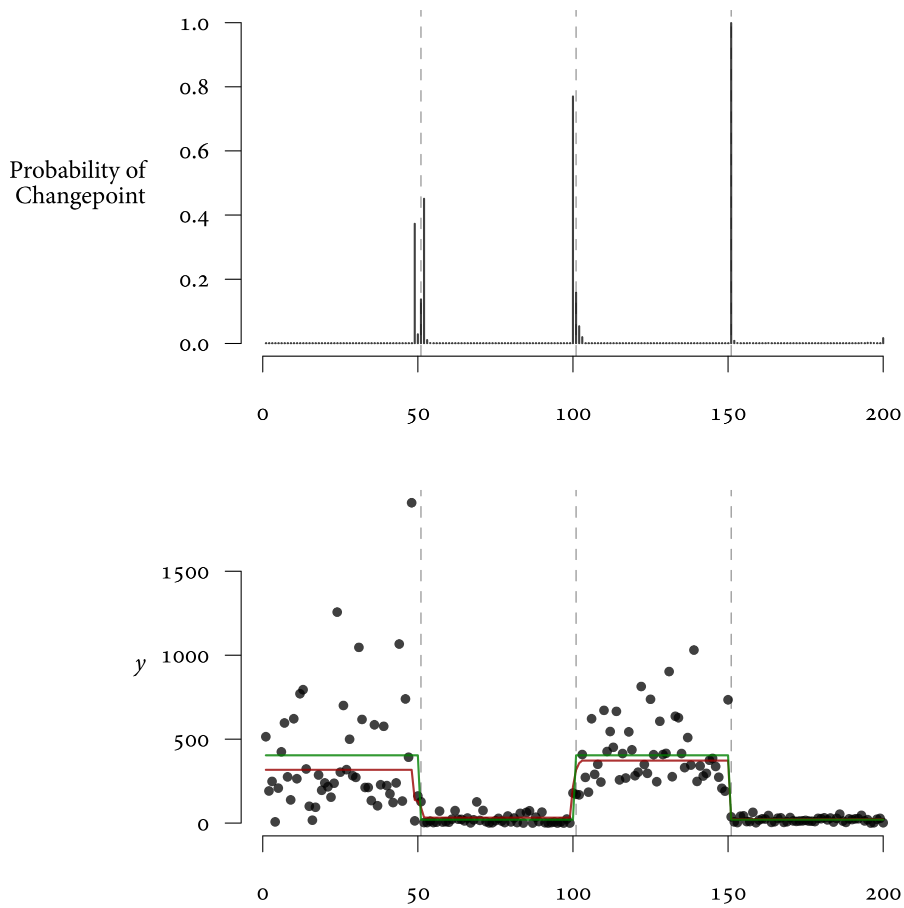


*Figure 2:* Changepoints in a simulated example. On top, the posterior probability of a changepoint in a given period. On bottom, the simulated data with the true daily mean (green), posterior mean (blue), and true changepoints (vertical and dashed).

regimes. Regime 2 in one draw could be very different from regime 2 in another draw. Instead of investigating the posterior mean of the regime parameters, we can estimate the posterior mean of the observation. That is, we can estimate

$$\hat{\lambda}_t = \frac{1}{G} \sum_{g=1}^{G} \exp(X_t \hat{\beta}^{(g)}_{s_t^{(g)}}), \tag{20}$$

where $G$ is the number of MCMC draws, $\hat{\beta}_k^{(g)}$ is the draw of $\beta_k$ in iteration $g$ of the sampler. The bottom panel of Firgure 2 overlays the true values of $\lambda_t$ in green along with its posterior mean, $\hat{\lambda}_t$ in red. In this case, the posterior values largely matched up the truth, with some (small) shrinkage toward the prior.

Extant changepoint models in political science also rely on the Poisson distribution, but do not take into account overdispersion. To demonstrate this, I applied 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. The top panel of Figure 3 shows that none of the estimated changepoints come close to the true changepoints. The bottom panel of the same figure shows why the Poisson model fails to find these changepoints. This panel plots a posterior predictive check (Gelman et al. 2003) 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.

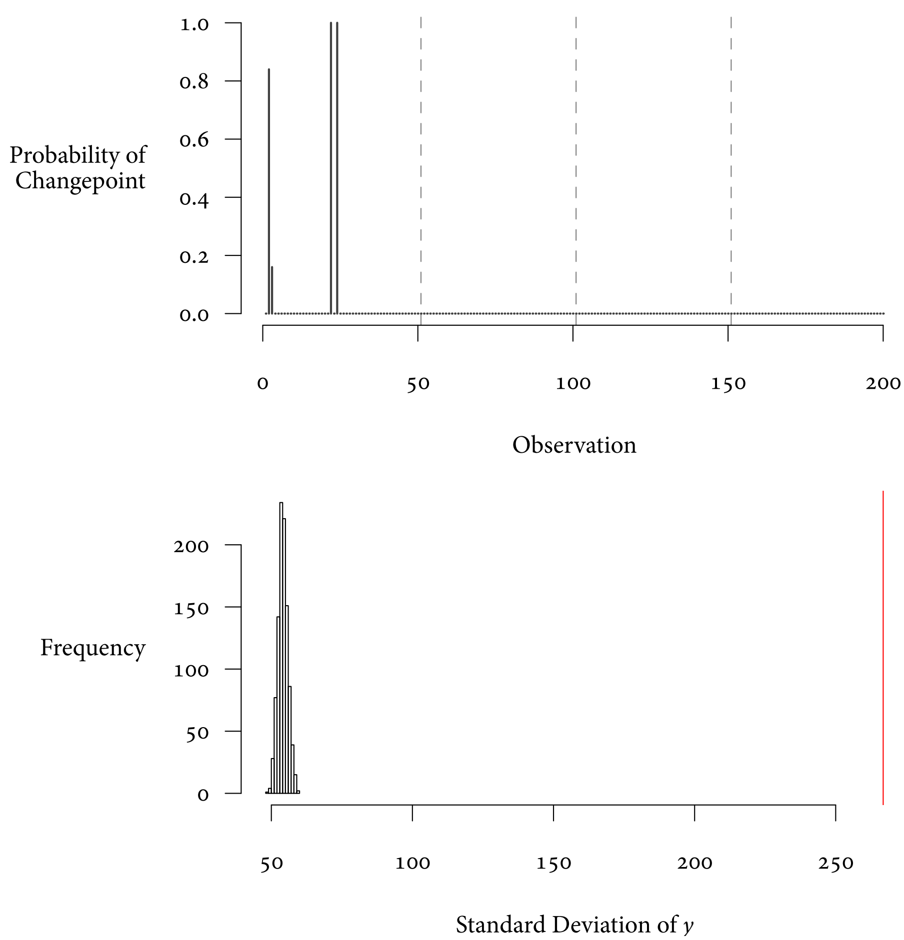


*Figure 3:* The results of a Poisson changepoint model with the simulated data. The top panel plots the posterior probability of a changepoint, with the true changepoints in dashed vertical lines. The bottom panel shows a posterior predictive check on the Poisson model, with a histogram of the standard deviations predicted by the posterior and the true standard deviation of the data in red.

| Estimated changepoint | Pr(Change) | Direction | Campaign Event |
|---|---|---|---|
| May 6, 2011 | 0.892 | + | Fox News debate (May 5) |
| May 21–25, 2011 | 0.945 | + | Announces candidacy (May 21) |
| September 23–25, 2011 | 0.784 | + | Wins Florida 5 Straw Poll (Sept. 24) |
| November 11, 2011 | 0.562 | − | Sexual misconduct allegations (Nov 7) |
| December 3, 2011 | 0.851 | − | Suspends campaign (Dec. 3) |

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

### *4.2 The rise and fall of Herman Cain*

Herman Cain's campaign for the 2012 Republican Presidential nomination provides an excellent demonstration of the validity of the above model. Cain was one of many candidates vying for the nomination and one of a few to reach the status of frontrunner, then quickly losing that status due in part to allegations of sexual misconduct. The ups and downs of Cain's campaign provide a good target for the changepoint model.

To estimate this model, I use the above MCMC sampler with 100,000 iterations, thinned by 100, with a burn-in period of 5,000 iterations. Figure 4 presents the posterior probability of a changepoint in the top panel. In the bottom panel, I plot the raw number of contributors along with the posterior mean of $\lambda_t$, 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 campaign event in the campaign.

Although Cain officially announced his candidacy on May 21, 2011, he did participate in campaign activities before that time, including a Fox News debate on May 5th, where at least one Fox News focus group voted him the "winner." The model predicts a changepoint the day after this debate along with a short regime of high activity after he officially announces his candidacy. The model then estimates a long summer regime of June until late September when the model finds a series of changepoints following Cain's winning of the Florida 5 Straw Poll (Sutton and Holland 2011). This

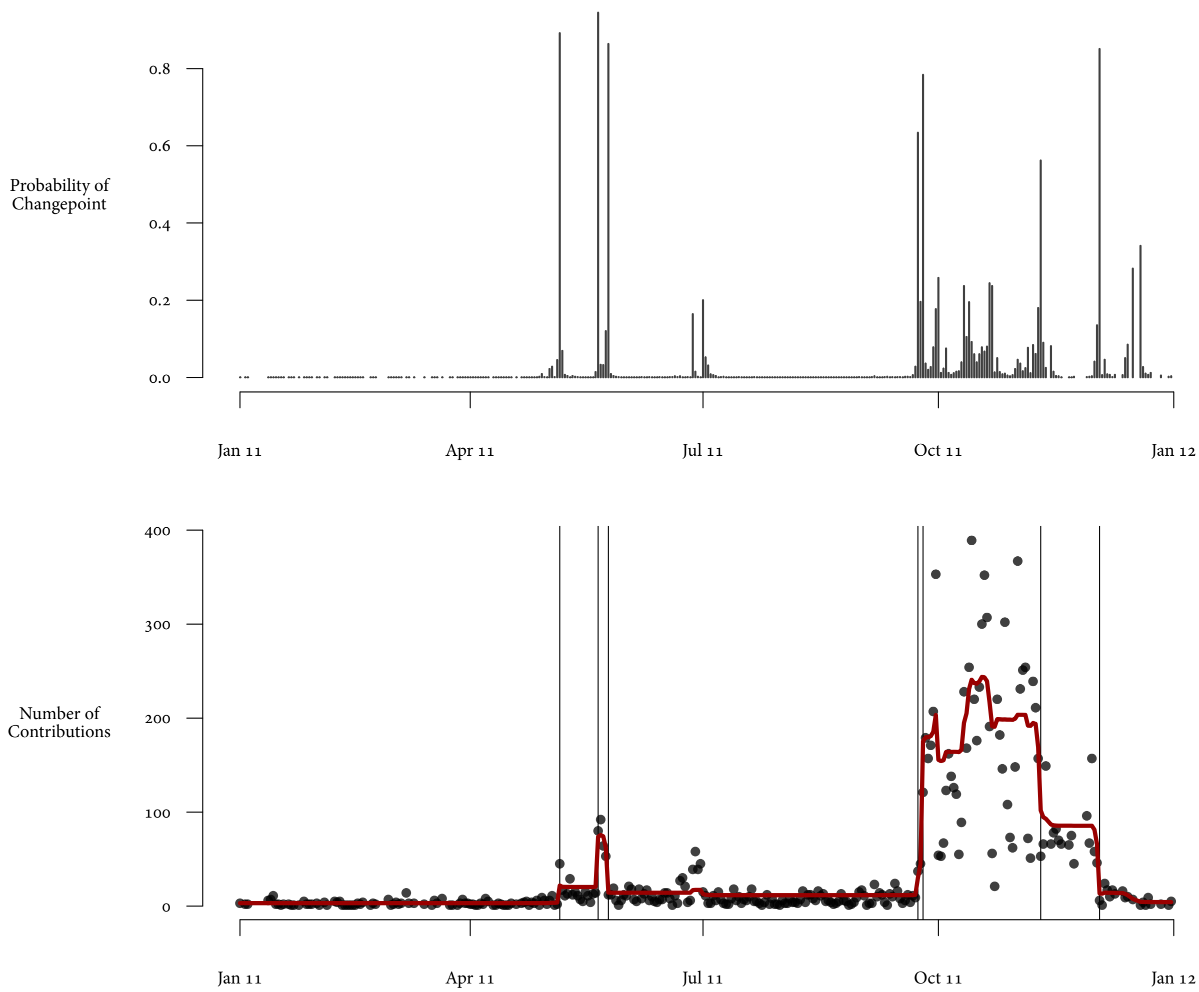


*Figure 4:* Contributions and changepoints for Herman Cain in the 2012 Republican Primary

regime of increased contributions lasts a little over a month until November 10th, a little over a week after the first reports of Cain's sexual misconduct on October 30th (Martin et al. 2011) and a few days after the first women to go public with accusations against Cain on November 7th (Henderson 2011). This decidedly lower regime is ended by an estimated changepoint on the day that Cain suspends his campaign for the nomination.

The model correctly identify major shifts in the distribution of contributions to Herman Cain which correspond to actual prominent events in his campaign. It is important to note that the model makes no restrictions on the number of changepoints in the data. This is crucial in this example, because it is difficult to specify the number of changepoints *a priori*, even if one were to visually inspect the time series. Furthermore, a Poisson changepoint model using the estimated number of changepoints (seven) from this output, fails to recover substantively important breaks. For example, the Poisson model fails to find a changepoint after the allegations of sexual harassment in early November. If we used this approach to help investigate the relationship between scandal and contributions, the Poisson model would lead us badly astray. The game-changers model ignores small blips in the data due to overdispersion and captures meaningful changes to the distribution of contributions.

### *4.3 The senatorial surges*

We can fruitfully apply this changepoint model to campaigns other than those at the national level. Furthermore, investigating local campaigns can give us insight into the relationship between local and national politics. To demonstrate this, I applied the gamechangers model to the fundraising for major-party nominees for Senate in the 2007-2008 election cycle.[3] One approach to modeling multiple candidates at the same time would be to build a hierarchical version of the gamechangers model and run this larger model on all the candidates at the same time. This approach is slightly problematic for a changepoint problem. Namely, there is no reason, *a priori*, to think that the regimes

3. Due to small sample sizes, I dropped any candidate that had fewer than 200 contributors or fewer than 100 days of positive contributions. This left 59 out of a possible 68 candidates in 34 races. Note that there were two states, Mississippi and Wyoming, who had two Senate elections in 2008 due to special elections to replace vacated seats.

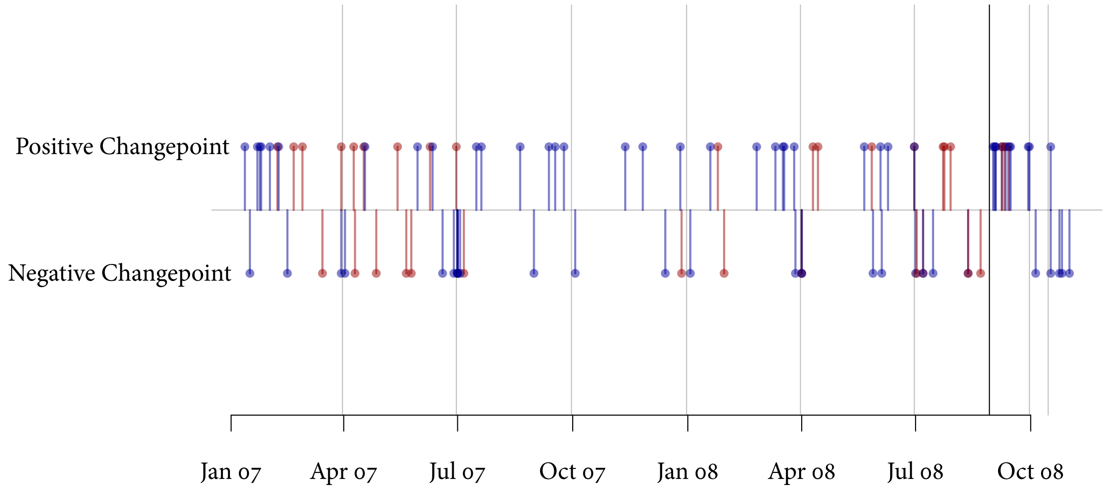


*Figure 5:* Changepoints for Senate races in 2008. The light grey lines are FEC filing deadlines. The black vertical line is the end of the Democratic National Convention.

of one campaign are necessarily the same fundamental type as regimes from another campaign. That is, it makes no sense to use the parameters from regime 2 in one race to help estimate the parameters in regime 2 in another race, since (a) regime 2 might be 1 day in one race and 100 days in the other or (b) regime 2 might not even occur in one of the races. To avoid this, we would have to either fix the number of changepoint across campaigns or implement a fairly complicated prior structure. Instead, I take the conceptually simpler approach and run the gamechangers model separately on each campaign. As in Section 4.2, I draw 100,000 MCMC iterations, thinned by 100, after an initial burn-in period of 5,000 draws.

The results from these models are presented in Figure 5, with the dates of Democratic changepoints in blue and Republican changepoints in red. In addition, the figure indicate the direction of the changepoint depending on the sign of $\hat{\lambda}_t - \hat{\lambda}_{t-1}$, where $t$ is the changepoint and $\hat{\lambda}_t$ is the mean of the posterior mean of the negative binomial at time $t$. The broad strokes of these results present an interesting picture. There is a flurry of activity early in the election cycle, then a relative calm in late 2007, then a steady pace in 2008. It is interesting to note that incumbent candidates dominate the "early money" gamechangers: 43 of the 48 changepoints in 2007 are for incumbent candidates (the changepoints in 2008 are almost exactly evenly divided between incumbents and non-incumbents).

In addition to locating the changepoints for each race, the game-changers model allows us to identify candidates who have certain type of changepoints. For instance, we may be interested in "surging" candidates: those whose fundraising takes off toward the end of the race. It is useful to identify these candidates, because they may give us insight as to how elites choose to contribute in close races.

| State | Candidate | Changepoint | % vote | CQ (Spring) | CQ (Fall) |
|---|---|---|---|---|---|
| AK | Begich (D) | Sep. 29, 2008 | 50.66 | Lean R | Leans D |
| CO | Udall (D) | Sep. 01, 2008 | 55.41 | Tossup | Leans D |
| MN | Franken (D) | Sep. 03, 2008 | 50.01 | Tossup | Tossup |
| NC | Hagan (D) | Sep. 11, 2008 | 54.37 | Likely R | Tossup |
| NH | Shaheen (D) | Sep. 08, 2008 | 53.27 | Tossup | Likely D |
| OR | Merkley (D) | Sep. 03, 2008 | 51.77 | Lean R | Tossup |

*Table 2:* Senate candidates who surged in 2008, as determined by the changepoint model. The "% vote" column is the their share of the two-party vote on election day. The CQ scores are the predictions made by *Congressional Quarterly* about the race in the Spring and the Fall.

To identify the surgers, I find all the campaign that had changepoints from September 1st, 2008 onward and that reached their maximum average contributions in the two months before election day. Table 2 shows the campaigns that meet this criteria in 2008, along with predictions from *Congressional Quarterly* and the final election outcome. Of these, five candidates are Democrats facing Republican incumbents, with only Rep. Mark Udall (CO) running for an open seat. All of the surging candidates had Spring predictions were either tossups or favoring the Republican. By October, the CQ rating had either remained the same or now favored the Democrat in each of the races. Furthermore, each of these candidates ended up wining their race, albeit sometimes by small margins.

Interestingly, all of these candidates were identified by various Democratic fundraising groups as being targets for overturning Republican-held seats. During the month of September, former Vice-President Al Gore sent emails to members of the liberal group MoveOn.org to encourage them to donate to the campaigns of Hagan, Franken, and Udall (Davis 2008). Early in September, a group of prominent Hollywood women organized a group called "Voices for a Senate Majority" which sought to raise at least $100,000 for each of these candidates (Ressner 2008). The ability of candidates to raise funds is often thought of as critical and investigating why and how certain candidates are able to surge in such contribution toward the end of the race could bring valuable insights into the causes and consequences of campaign contributions more broadly. The game-changers models allows this kind of study by identifying these surging campaigns. The measurement of these surges would be key to a study of how external actors (like MoveOn) can influence a candidate's rise during the election season.

### 4.4 *Game-changers and news coverage*

Now that we have estimated changepoints for each Senate candidates for 2008, we may wish to understand what relationship these game-changers have with other aspects of the campaign. One way in which periods with changepoints differ from periods without changepoints is in the how the press covers the them. To demonstrate this, I collected data on the amount of coverage dedicated to each Senate race in each week using a political trade publication called *The Bulletin's Frontrunner*. This daily publications provides summaries of the national and local news coverage of each race. To measure the amount of coverage, I count the number of words in these summaries aggregated up the weekly level. This measure varies from zero words in some weeks to up to roughly 3,000 words toward the end of the campaign.

To get a sense for how game-changers relate to news coverage, I ran a logistic regression of the presence of a changepoint in a given race in a given week on the number of words written about that race in the *Frontrunner*. In addition, I included a linear time trend, the number of ads run by the candidates or the parties in that week, the Democratic percent of two-party poll results in that week, and, in some specifications, a race fixed-effect. Figure 6 shows how the probability of a changepoint

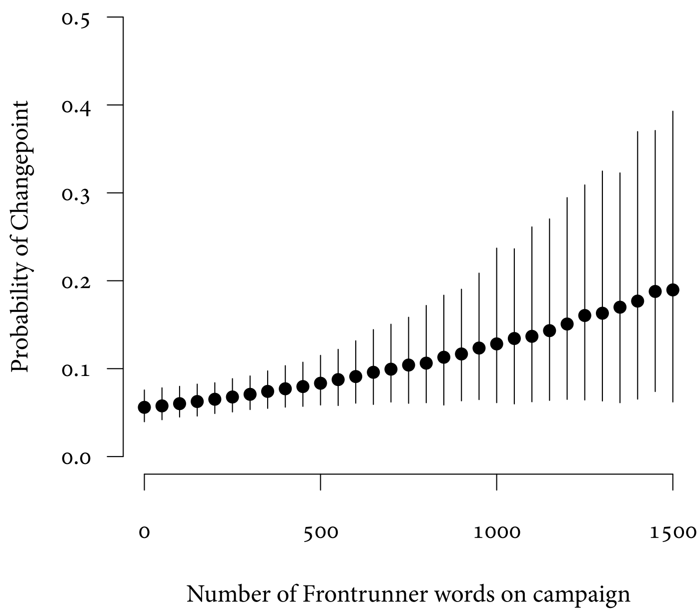


*Figure 6:* Probability of a changepoint as a function of the news coverage as measured by the *Frontrunner* word count.

changes with the *Frontrunner* word count.[4] In weeks with more news coverage, there is a greater chance of a changepoint and moving from 0 words to 1000 words roughly double those chances. This result seems to indicate that the dynamics of contributions might depend heavily on the attention paid to a campaign.

## §5 Conclusion

Some campaigns take off and some campaigns fall flat. This paper presents a novel statistical model that estimates the number and timing of these changepoints in campaign contributions data. This model gives researchers the ability to detect significant events in campaigns and investigate the nature of these shifts in the broader political context. This represents the first attempt to measure a fairly tricky, yet common phenomenon: a campaign game-changer. With the game-changers model in hand, we can estimate changepoints for a whole host of campaigns and for a whole host contribution types—individuals versus PACs, men versus women, or in-state versus out-of-state. Further exploring the variation in structural breaks will help us better understand the nature of contributions as political participation.

Methodologically, the game-changer model pushes changepoint models forward by bringing together a few novel features. First, it naturally incorporates the overdispersion that is common in count data. Second, it leans on Bayesian nonparametrics in order to estimate the number of changepoint instead of having to know it *a priori*. This second contribution is especially important when, as in this case, marginal likelihoods are difficult to compute. One obvious way to extend this model is to build a multivariate version of the game-changers model. This model would estimate the changepoints for multiple time-series at the same time, allowing for in-model comparisons and complicated dependence structures. A potentially useful approach might be to combine the present model with the dynamic overdispersion model of Brandt and Sandler (2012).

4. This was generated by the simulation-based marginal effects method of King, Tomz, and Wittenberg (2000), using `Zelig` (Imai, King, and Lau 2006).

I have tailored the model to campaign contributions data, but the applications of this model reach far beyond campaigns. In international relations, the number of violent attacks or deaths in a conflict are likely to feature overdispersion due to geographic and strategic clustering. The present model could overcome this issue and help scholars identify conflict regimes during the course of a conflict. In many areas of the social sciences, scholars engage in *event studies* to identify the patterns that underlie how certain events arise. Scholars facing overdispersion in their event study could fruitfully apply this model to their specific problem.

The Dirichlet process prior approach that I take in this paper is more general than this specific negative binomial outcome model. Because it generalizes the Chib (1998) method for multiple changepoints, it also inherits the broad applicability of that method. Since the model parameters $\Theta$ are drawn conditional on the latent states and the Dirichlet process prior only affects the drawing of the latent state, it is straightforward to adapt this approach to changepoint model for continuous, binary, and ordered categorical variables such as those in Park (2011) or Spirling (2007).

## References


Bartels, Larry M. 1985. "Expectations and preferences in presidential nominating campaigns." *American Political Science Review*:804–815. (Cit. on p. 1).

Brandt, Patrick T., and Todd Sandler. 2012. "A Bayesian Poisson Vector Autoregression Model." *Political Analysis.* (Cit. on p. 25).

Chib, Siddhartha. 1995. "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* 90 (432): pp. 1313–1321. (Cit. on p. 8).

———. 1998. "Estimation and comparison of multiple change-point models." *Journal of Econometrics* 86 (2): 221–241. (Cit. on pp. 2, 5, 6, 8, 10, 12, 13, 26).

Chib, Siddhartha, and Edward Greenberg. 1996. "Markov Chain Monte Carlo Simulation Methods in Econometrics." *Econometric Theory* 12 (3): pp. 409–431. (Cit. on p. 5).

Chong, Terence Tai-Leung, and Stanley Iat-Meng Ko. 2011. "Dirichlet Process Multiple Change-point Model." (Cit. on p. 8).

Christenson, Dino P, and Corwin D Smidt. 2011. "Riding the Waves of Money: Contribution Dynamics in the 2008 Presidential Nomination Campaign." *Journal of Political Marketing* 10, nos. 1-2 (February): 4–26. (Cit. on p. 4).

Davis, Susan. 2008. "Gore, MoveOn Team Up to Aid Senate Candidates." *Wall Street Journal* (September 7). (Cit. on p. 23).

Escobar, Michael D., and Mike West. 1995. "Bayesian Density Estimation and Inference Using Mixtures." *Journal of the American Statistical Association* 90 (430): pp. 577–588. (Cit. on p. 9).

Federal Election Commission. 2011. *Campaign Guide for Congressional Candidates and Committees.* (Cit. on p. 3).

Ferguson, Thomas S. 1973. "A Bayesian Analysis of Some Nonparametric Problems." *The Annals of Statistics* 1 (2): pp. 209–230. (Cit. on p. 9).

Frühwirth-Schnatter, Sylvia, Rudolf Frühwirth, Leonhard Held, and Håvard Rue. 2009. "Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data." *Statistics and Computing* 19 (4): 479–492. (Cit. on pp. 6, 11, 13).

Gelman, Andrew, J.B. Carlin, H.S. Stern, and D.B. Rubin. 2003. *Bayesian Data Analysis, Second Edition.* Chapman & Hall. (Cit. on p. 16).

Geweke, John, and Yu Jiang. 2011. "Inference and prediction in a multiple-structural-break model." *Journal of Econometrics* 163 (2): 172–185. (Cit. on p. 8).

Giordani, Paolo, and Robert Kohn. 2008. "Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models." *Journal of Business & Economic Statistics* 26, no. 1 (January): 66–77. (Cit. on p. 8).

Grimmer, Justin. 2011. "An Introduction to Bayesian Inference via Variational Approximations." *Political Analysis* 19 (1): 32–47. (Cit. on p. 9).

Henderson, Nia-Malika. 2011. "Sharon Bialek accuses Herman Cain of sexual harassment as she sought help getting a job." *Washington Post* (November 7). (Cit. on p. 20).

Imai, Kosuke, Gary King, and Olivia Lau. 2006. "Zelig: Everyone's Statistical Software." (Cit. on p. 25).

Ishwaran, Hemant, and Lancelot F. James. 2001. "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association* 96 (453): 161–173. (Cit. on p. 10).

King, Gary. 1989. "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33, no. 3 (August): 762–784. (Cit. on pp. 2, 7).

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44, no. 2 (April): 341–355. (Cit. on p. 25).

Koop, Gary, and Simon M. Potter. 2007. "Estimation and forecasting in models with multiple breaks." *Review of Economic Studies* 74 (3): 763–789. (Cit. on p. 8).

———. 2009. "Prior Elicitation in Multiple Change-point Models." *International Economic Review* 50 (3): 751–772. (Cit. on p. 8).

Martin, Jonathan, Maggie Haberman, Anna Palmer, and Kenneth P. Vogel. 2011. "Herman Cain accused by two women of inappropriate behavior." *Politico* (October 31). (Cit. on p. 20).

Mutz, Diana C. 1995. "Effects of Horse-Race Coverage on Campaign Coffers: Strategic Contributing in Presidential Primaries." *The Journal of Politics* 57 (04): 1015–1042. (Cit. on p. 1).

Neal, Radford M. 2000. "Markov Chain Sampling Methods for Dirichlet Process Mixture Models." *Journal of Computational and Graphical Statistics* 9 (2): pp. 249–265. (Cit. on pp. 9, 10).

Park, Jong Hee. 2010. "Structural Change in U.S. Presidents' Use of Force." *American Journal of Political Science* 54 (3): 766–782. (Cit. on pp. 2, 6, 8, 16).

———. 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. (Cit. on pp. 6, 8, 26).

Ressner, Jeffrey. 2008. "Smile! You're in PoliWood." *Politico* (September 4). (Cit. on p. 23).

Spirling, Arthur. 2007. "Bayesian Approaches for Limited Dependent Variable Change Point Problems." *Political Analysis* 15 (4): 387–405. (Cit. on pp. 2, 6, 26).

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): 447–457. (Cit. on p. 9).

Sutton, Jane, and Steve Holland. 2011. "Cain upsets Perry in Florida Republican straw poll." *Reuters* (September 24). (Cit. on p. 18).

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Volunteerism in American Politics.* Cambridge, MA: Harvard University Press. (Cit. on p. 1).