Game-changers:

Detecting shifts in the flow of campaign contributions

Matthew Blackwell
University of Rochester

March 8th, 2013

APWG

Why not polls?

Why not polls?

1. Lack of variation

Why not polls?

1. Lack of variation

2. Cheap talk

Why not polls?

1. Lack of variation
2. Cheap talk
3. Data (un)availability

NAME OF COMMITTEE (In Full)
Obama for America

A. Full Name (Last, First, Middle Initial)
Sharon Anderson

Mailing Address 1668 finwick dr

City	State	Zip Code
pfafftown	NC	27040-9043

FEC ID number of contributing federal political committee. C

Name of Employer	Occupation
The Norman Group	Intl Consultant

Receipt For: 2012
[X] Primary [ ] General
[ ] Other (specify)

Election Cycle-to-Date: 271.00

Transaction ID : C19176830
Date of Receipt: 08 / 12 / 2012

Amount of Each Receipt this Period: 11.00

B. Full Name (Last, First, Middle Initial)
Riaz Hussain

Mailing Address 540 N Webster Ave

City	State	Zip Code
Scranton	PA	18510

FEC ID number of contributing federal political committee. C

Name of Employer	Occupation
University of Scranton	Professor

Receipt For: 2012
[X] Primary [ ] General
[ ] Other (specify)

Election Cycle-to-Date: 225.00

Transaction ID : C20196560
Date of Receipt: 08 / 30 / 2012

Amount of Each Receipt this Period: 35.00

C. Full Name (Last, First, Middle Initial)
Dave Baird

Mailing Address 1376 Lincoln St

Transaction ID : C20090710
Date of Receipt

A measurement question

When do campaigns take off or fall flat?

A measurement question

When do campaigns take off or fall flat?

When do campaign contributions take off or fall flat?

A measurement question

When do campaigns take off or fall flat?

When do campaign contributions take off or fall flat?

Tools: Bayesian nonparametric model for overdispersed count data.

Why contributions?

Why contributions?

1. Lots of variation

Why contributions?

1. Lots of variation

2. Costly participation

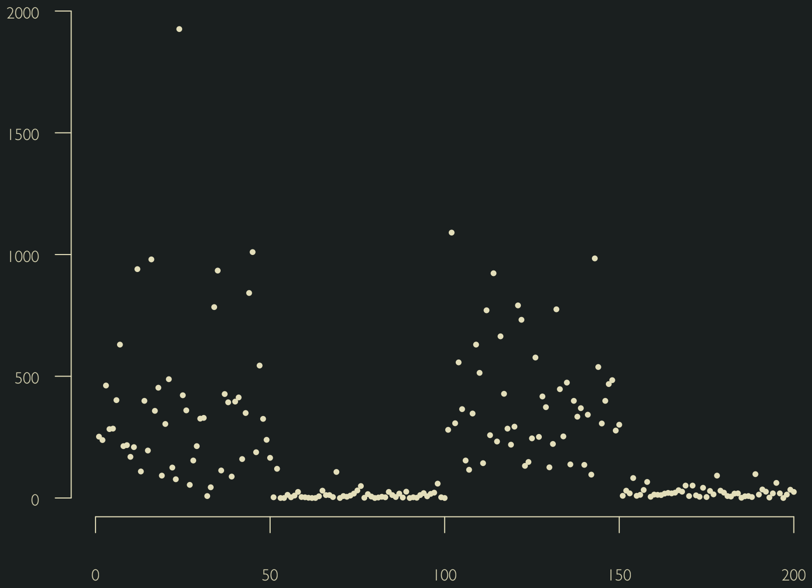
Why contributions?

1. Lots of variation

2. Costly participation

3. Data availability

Why changepoint models?



The challenges

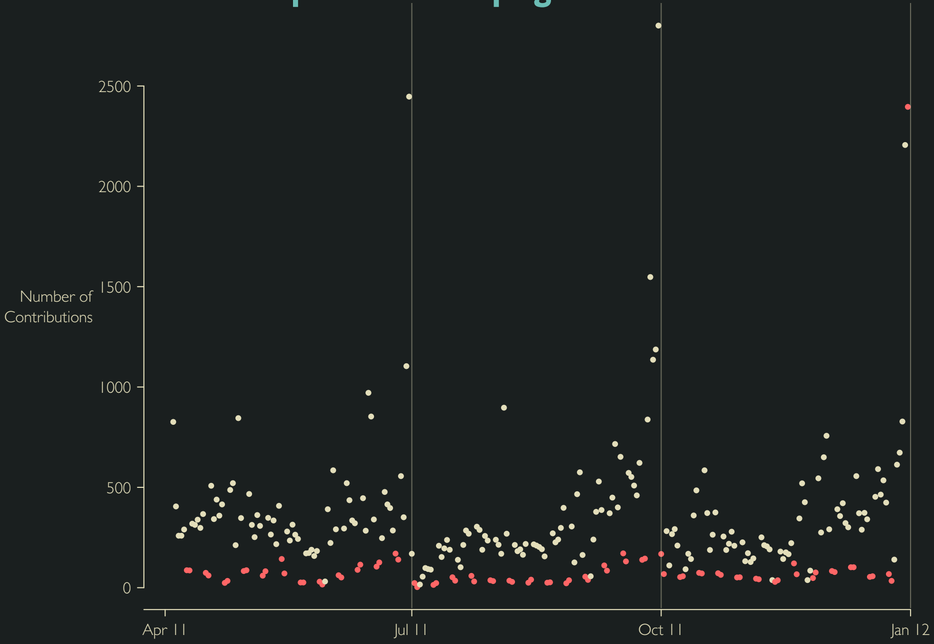
Modeling daily contribution counts

The challenges

Modeling daily contribution counts

Choosing the number of changepoints

Overdispersion in campaign contributions



Bayesian model for overdispersed counts

For observations t in $\{1, \ldots, T\}$:

$$[y_t | \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t) \qquad \text{(data)}$$

Bayesian model for overdispersed counts

For observations t in $\{1, \dots, T\}$:

$$[y_t | \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t) \qquad \text{(data)}$$

$$\lambda_t = \exp(X_t \beta) \qquad \text{(link function)}$$

Bayesian model for overdispersed counts

For observations t in $\{1, ..., T\}$:

$$[y_t | \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t) \qquad \text{(data)}$$

$$\lambda_t = \exp(X_t \beta) \qquad \text{(link function)}$$

$$[\eta_t | \rho] \sim \text{Gamma}(\rho, \rho) \qquad \text{(random effect)}$$

Bayesian model for overdispersed counts

For observations t in $\{1, \ldots, T\}$:

$$[y_t | \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t) \quad \text{(data)}$$

$$\lambda_t = \exp(X_t \beta) \quad \text{(link function)}$$

$$[\eta_t | \rho] \sim \text{Gamma}(\rho, \rho) \quad \text{(random effect)}$$

marginal distribution of y:

$$[y_t | \beta, \rho, X] \sim \text{NegBin}(\rho, \rho/(\rho + \lambda_t))$$

Generalize to a mixture model

$$[y_t | s_t, \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t) \qquad \text{(data)}$$

Generalize to a mixture model

$$[y_t | s_t, \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t) \qquad \text{(data)}$$

regimes $(1, \ldots, K)$ $\quad s_t = k$

Generalize to a mixture model

$$[y_t | s_t, \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t)$$ (data)

$$\lambda_t = \exp(X_t \beta_k)$$ (link function)

regimes $(1, \ldots, K)$ $s_t = k$

Generalize to a mixture model

$$[y_t | s_t, \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t)$$ (data)

$$\lambda_t = \exp(X_t \beta_k)$$ (link function)

regimes $(1, \ldots, K)$ $\quad s_t = k$

$$[\eta_t | \rho, s_t] \sim \text{Gamma}(\rho_k, \rho_k)$$ (random effect)

Generalize to a mixture model

$$[y_t | s_t, \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t)$$ (data)

$$\lambda_t = \exp(X_t \beta_k)$$ (link function)

regimes $(1, \dots, K)$ $\quad s_t = k$

$$[\eta_t | \rho, s_t] \sim \text{Gamma}(\rho_k, \rho_k)$$ (random effect)

$$\Pr(s_{t+1} = k \,|\, s_t = k) = p_k$$

Generalize to a mixture model

$$[y_t | s_t, \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t)$$ (data)

$$\lambda_t = \exp(X_t \beta_k)$$ (link function)

regimes $(1, \ldots, K)$ — $s_t = k$

$$[\eta_t | \rho, s_t] \sim \text{Gamma}(\rho_k, \rho_k)$$ (random effect)

$$\Pr(s_{t+1} = k \,|\, s_t = k) = p_k$$
$$\Pr(s_{t+1} = k + 1 \,|\, s_t = k) = 1 - p_k$$

Generalize to a mixture model

$$[y_t | s_t, \eta_t, \beta, \rho, X] \sim \text{Poisson}(\eta_t \lambda_t) \qquad \text{(data)}$$

regimes $(1, \ldots, K)$

$$s_t = k \;\rightarrow\; \lambda_t = \exp(X_t \beta_k) \qquad \text{(link function)}$$

$$s_t = k \;\rightarrow\; [\eta_t | \rho, s_t] \sim \text{Gamma}(\rho_k, \rho_k) \qquad \text{(random effect)}$$

$$\Pr(s_{t+1} = k \,|\, s_t = k) = p_k$$
$$\Pr(s_{t+1} = k+1 \,|\, s_t = k) = 1 - p_k$$
$$\Pr(s_{t+1} = j \,|\, s_t = k) = 0 \qquad (\forall j \notin \{k, k+1\})$$

Traditional changepoint models

Regimes

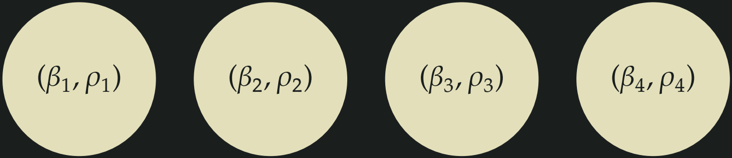
Traditional changepoint models

Regimes

(β_1, ρ_1) (β_2, ρ_2) (β_3, ρ_3) (β_4, ρ_4)

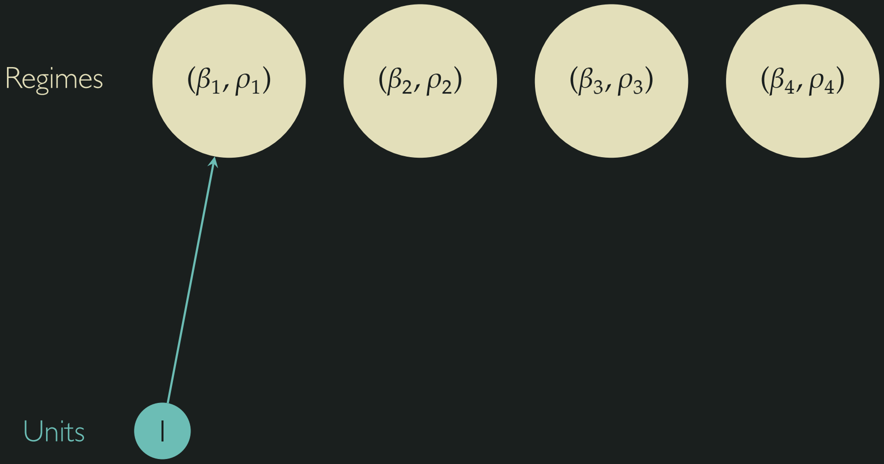
Traditional changepoint models

Regimes

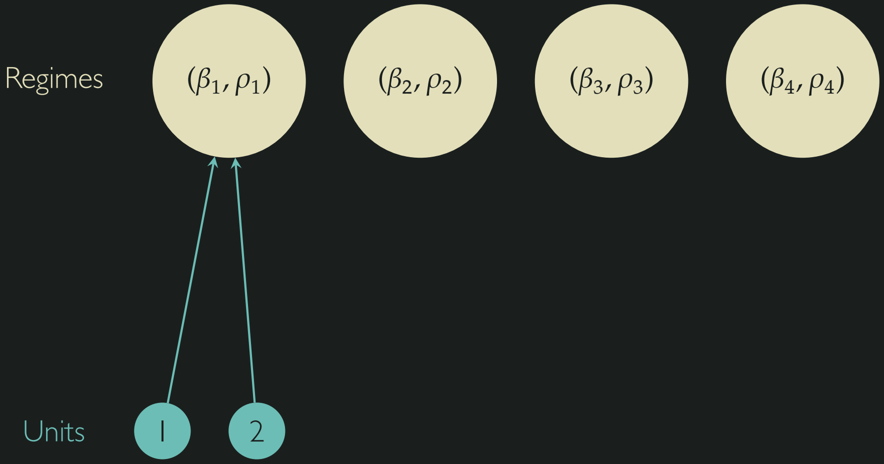


Units

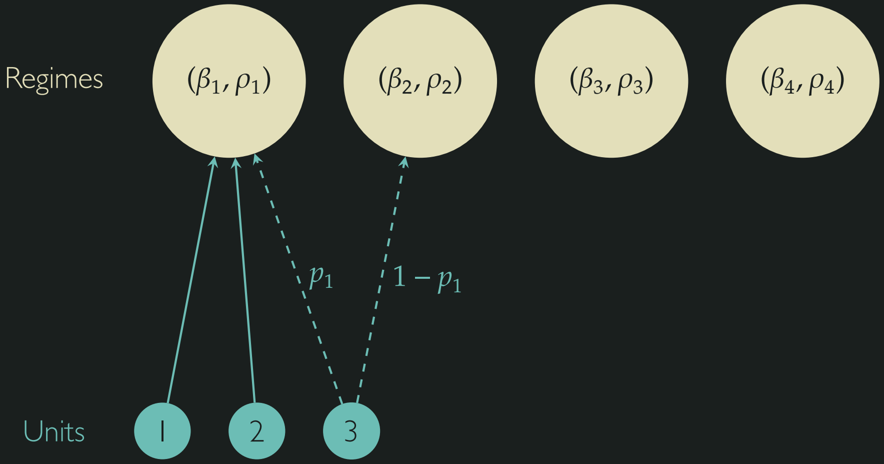
Traditional changepoint models



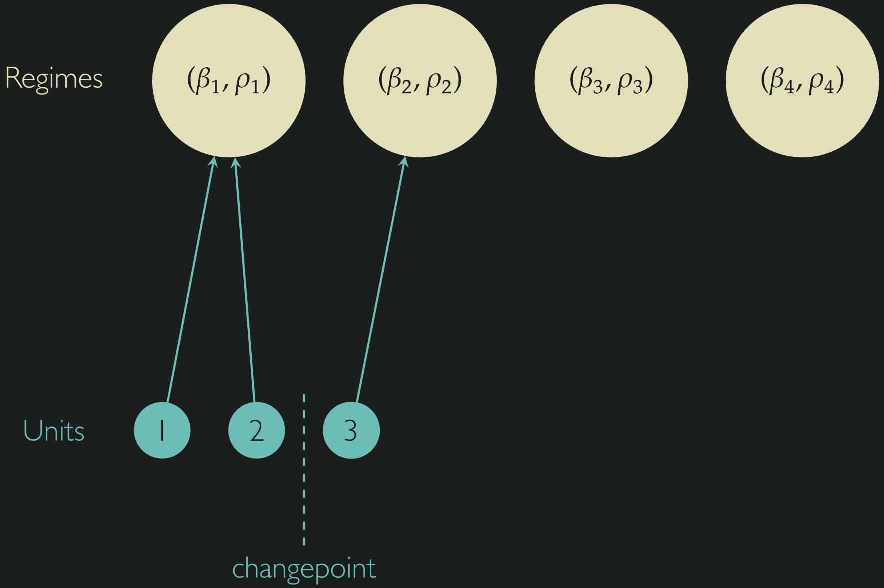
Traditional changepoint models



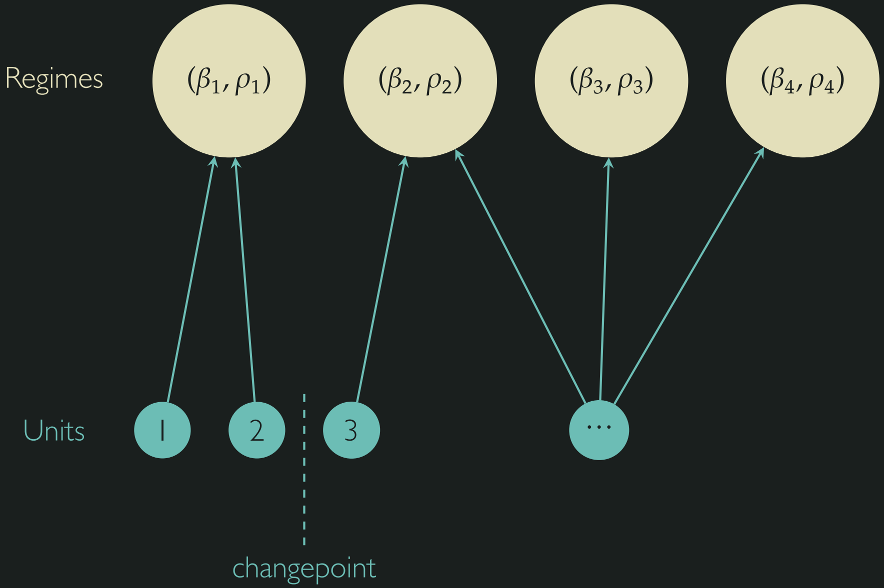
Traditional changepoint models



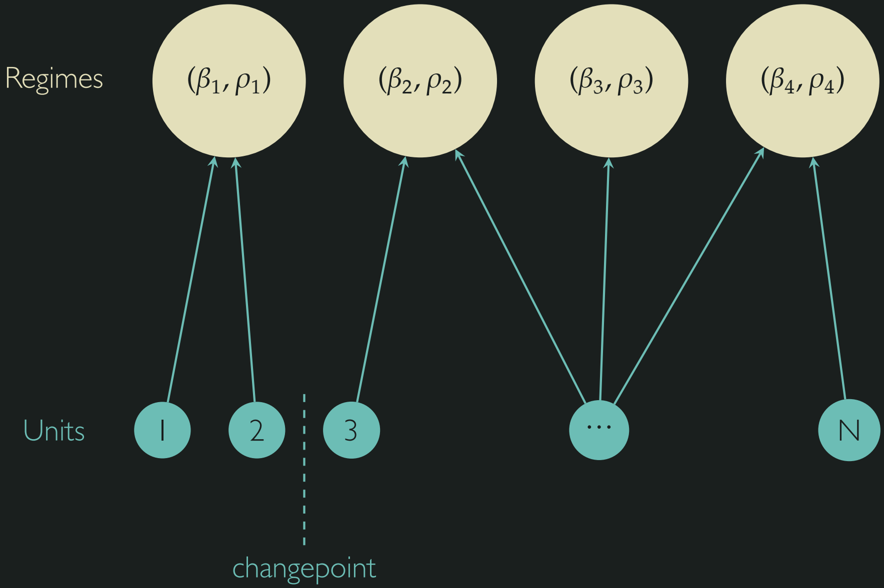
Traditional changepoint models



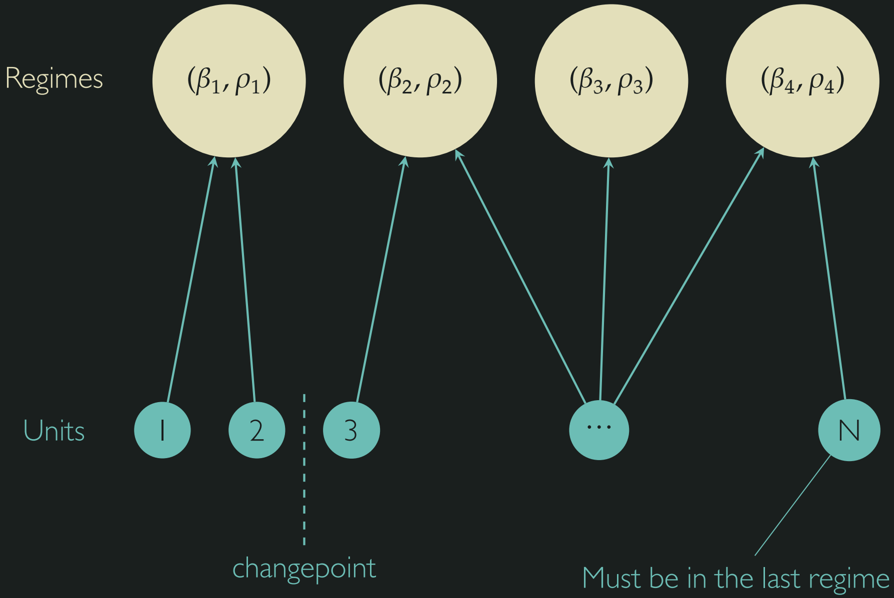
Traditional changepoint models



Traditional changepoint models



Traditional changepoint models



Bayesian nonparametric priors

- Model assumptions: $y_t \sim G$ i.i.d. from an unknown distribution G.

Bayesian nonparametric priors

- Model assumptions: $y_t \sim G$ i.i.d. from an unknown distribution G.
- Parametric structure on our priors puts limitations on the posterior inferences.

Bayesian nonparametric priors

- Model assumptions: $y_t \sim G$ i.i.d. from an unknown distribution G.

- Parametric structure on our priors puts limitations on the posterior inferences.

- Bayesian nonparametrics: priors over distributions and, thus, an infinite number of parameters.

Dirichlet process prior

Dirichlet process prior

- Dirichlet process prior clusters units into a countably infinite set of groups.

Dirichlet process prior

- Dirichlet process prior clusters units into a countably infinite set of groups.
- Obviously we only observe a finite number of these groups, but the number is determined by the data and the prior, not exclusively from the prior.

Dirichlet process prior

- Dirichlet process prior clusters units into a countably infinite set of groups.

- Obviously we only observe a finite number of these groups, but the number is determined by the data and the prior, not exclusively from the prior.

$$\text{regimes: } s_t \in \{1, \ldots, \infty\}$$

$$\Pr(s_{t+1} = k \mid s_t = k) = \frac{n_k}{t-1+b}$$

$$\Pr(s_{t+1} = k+1 \mid s_t = k) = \frac{b}{t-1+b}$$

Dirichlet process prior

Dirichlet process prior

Regimes

(infinite)

Dirichlet process prior

Regimes (infinite)

(β_1, ρ_1) (β_2, ρ_2) (β_3, ρ_3) $\cdots$ $\cdots$ $\cdots$

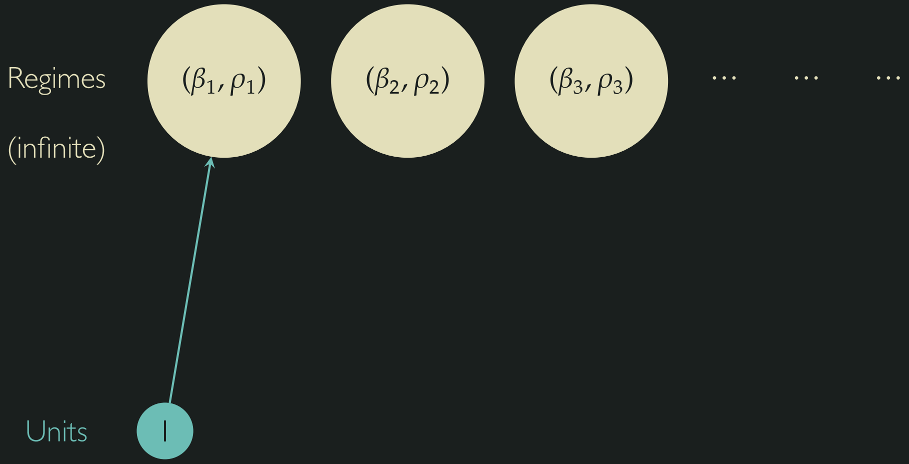
Dirichlet process prior

Regimes (infinite)

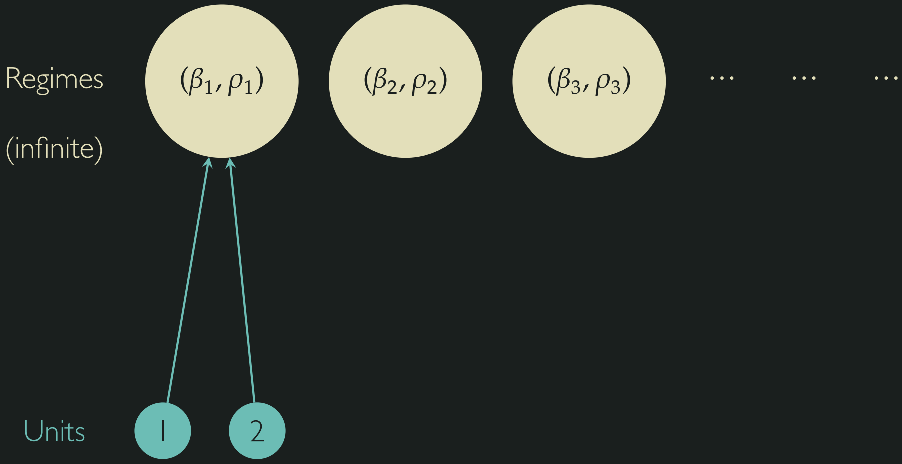
(β_1, ρ_1) (β_2, ρ_2) (β_3, ρ_3) $\cdots$ $\cdots$ $\cdots$

Units

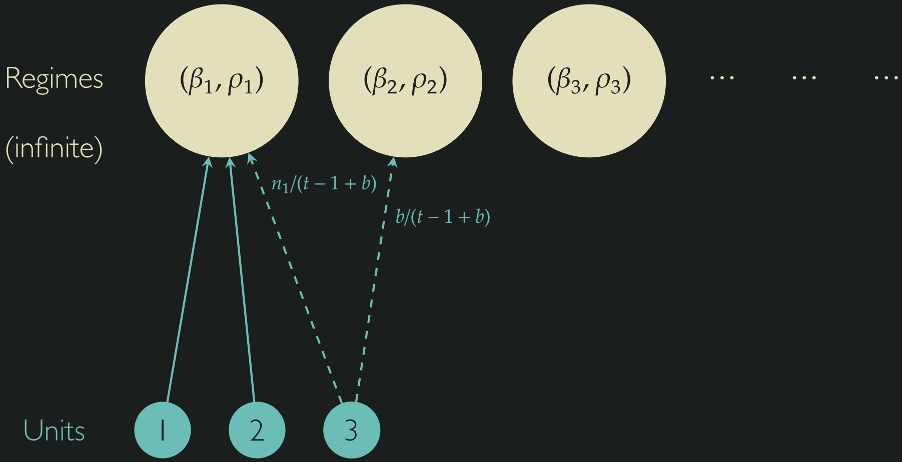
Dirichlet process prior



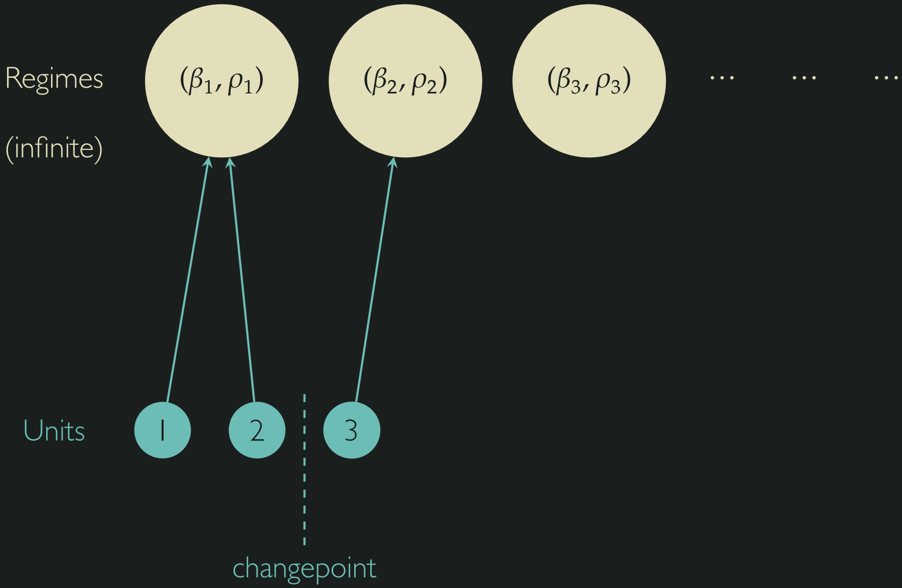
Dirichlet process prior



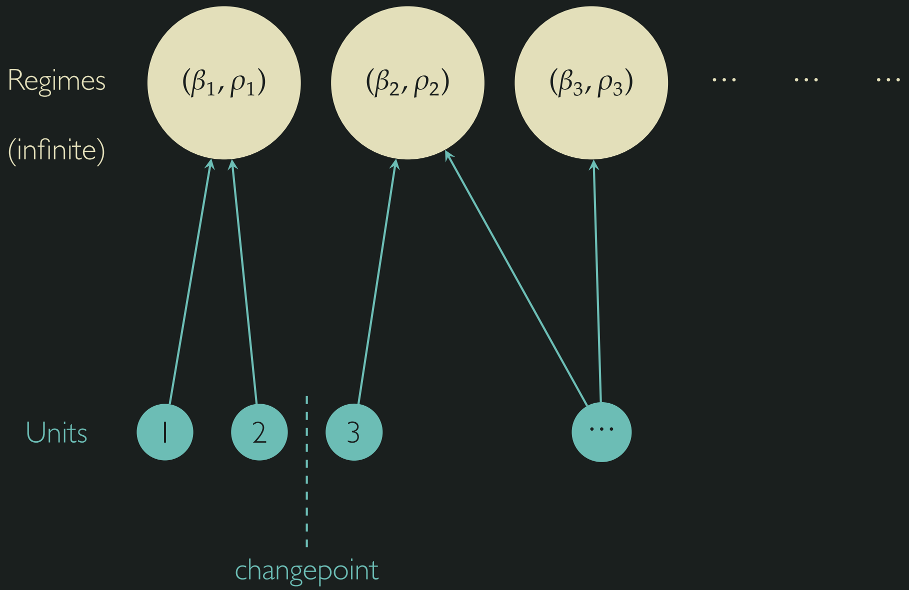
Dirichlet process prior



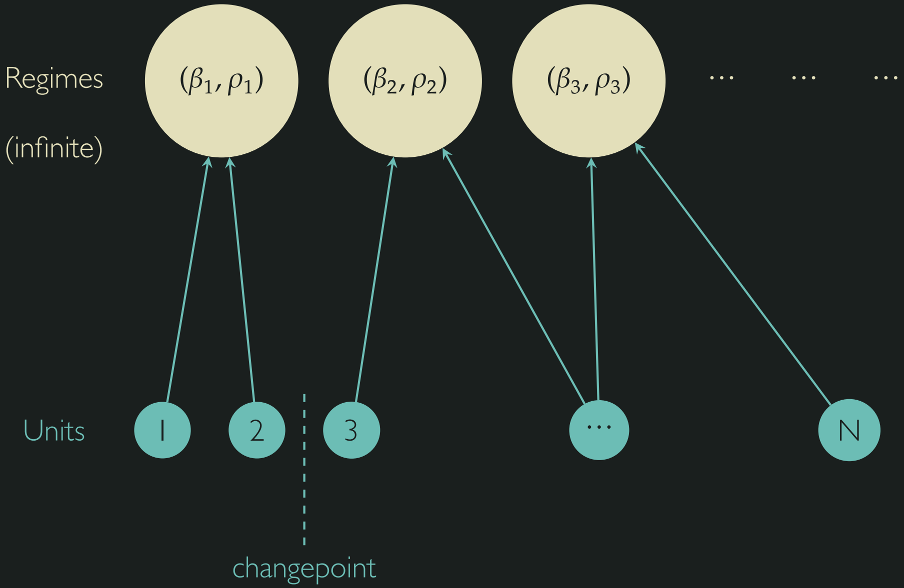
Dirichlet process prior



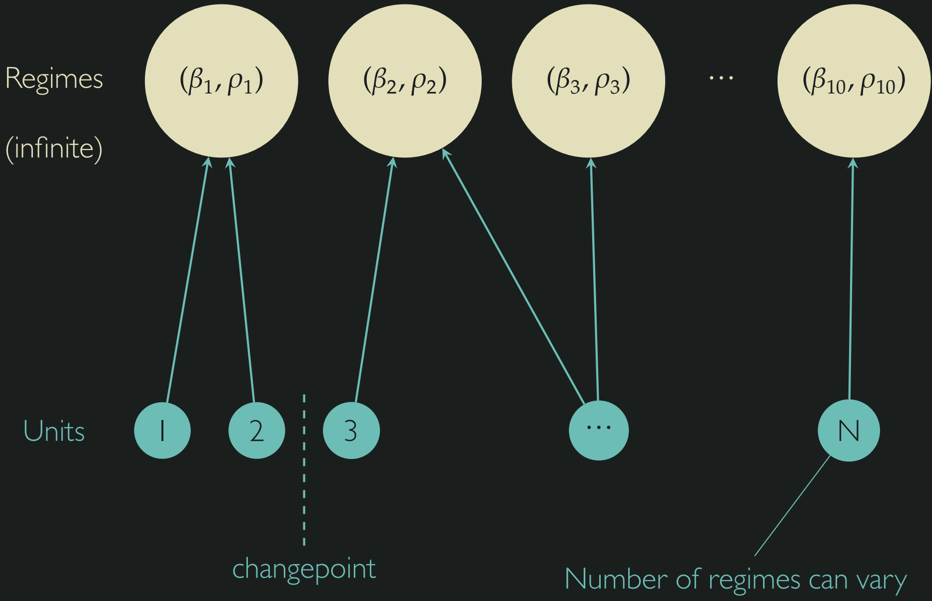
Dirichlet process prior



Dirichlet process prior



Dirichlet process prior



Your lunch is never free

- DPP has a rich-get-richer property:

$$\Pr(s_{t+1} = k \mid s_t = k) = \frac{n_k}{t - 1 + b}$$

Your lunch is never free

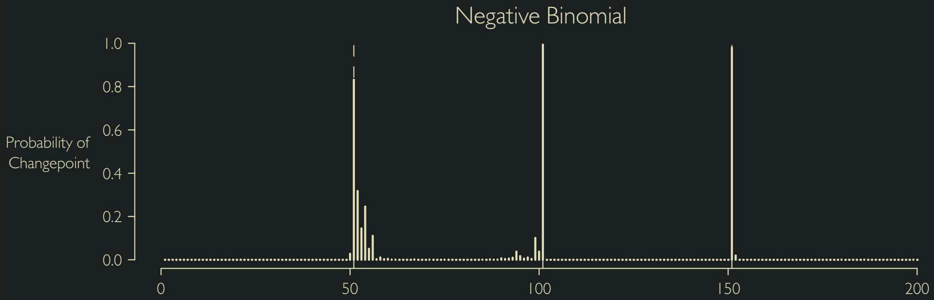
- DPP has a rich-get-richer property:

$$\Pr(s_{t+1} = k \mid s_t = k) = \frac{n_k}{t - 1 + b}$$

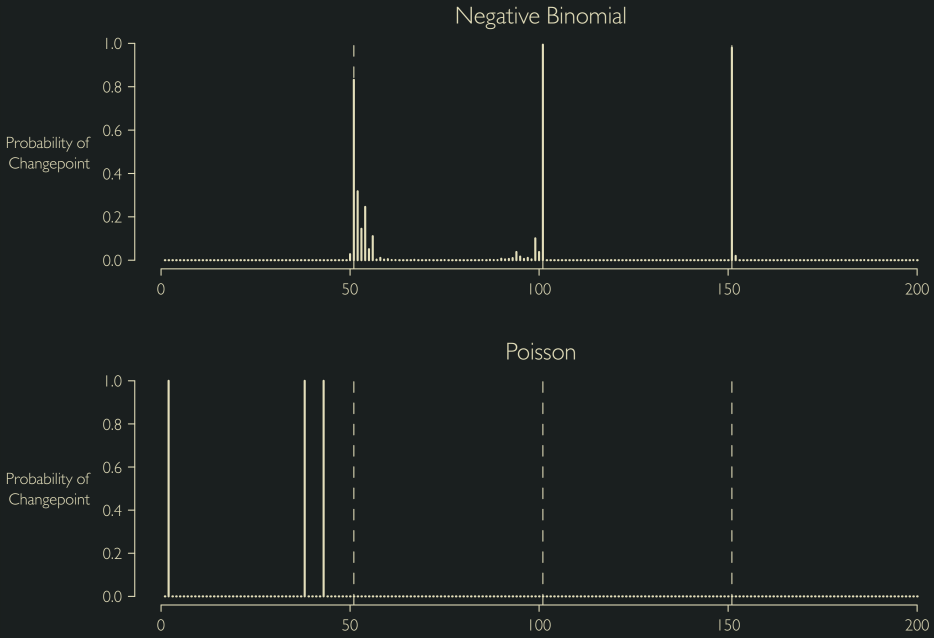
- No free lunch theorem: All nonparametric priors place assumptions on the clustering algorithm and no algorithm is optimal across the space of all problems.

Monte Carlo evidence

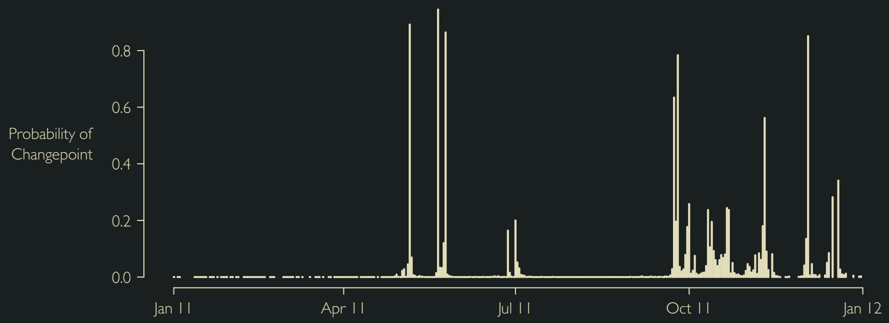
Monte Carlo evidence



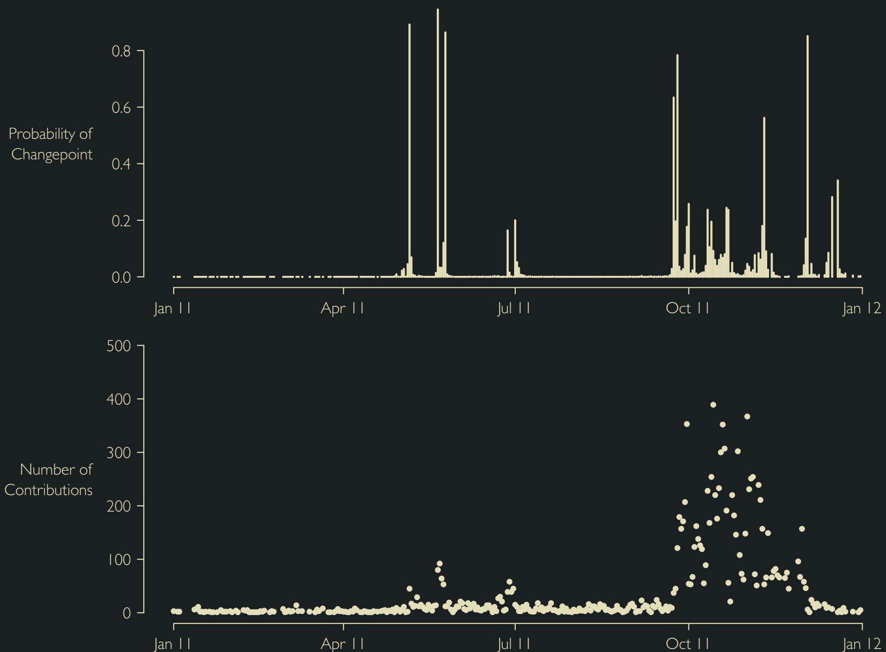
Monte Carlo evidence



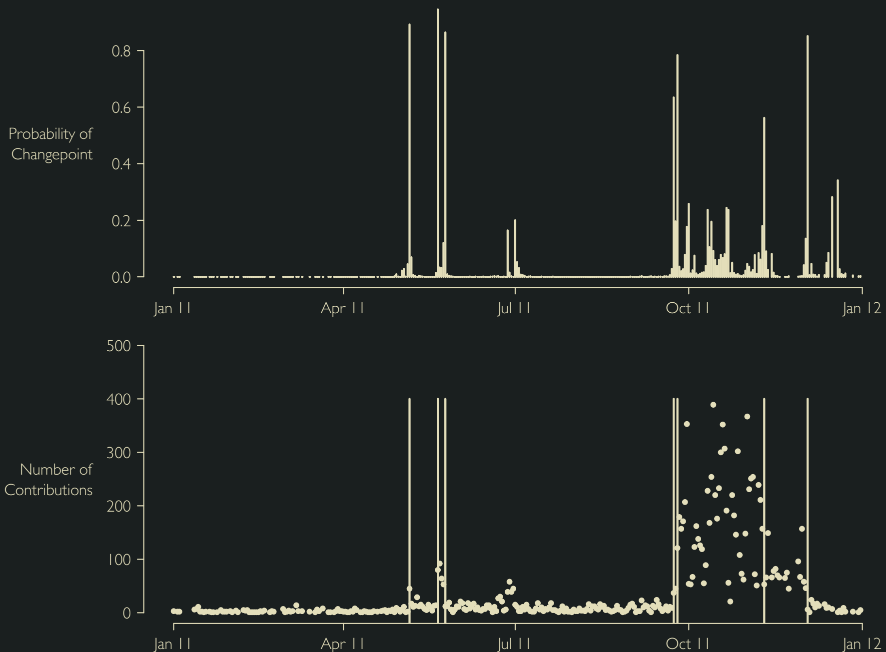
The rise and fall of Herman Cain



The rise and fall of Herman Cain



The rise and fall of Herman Cain

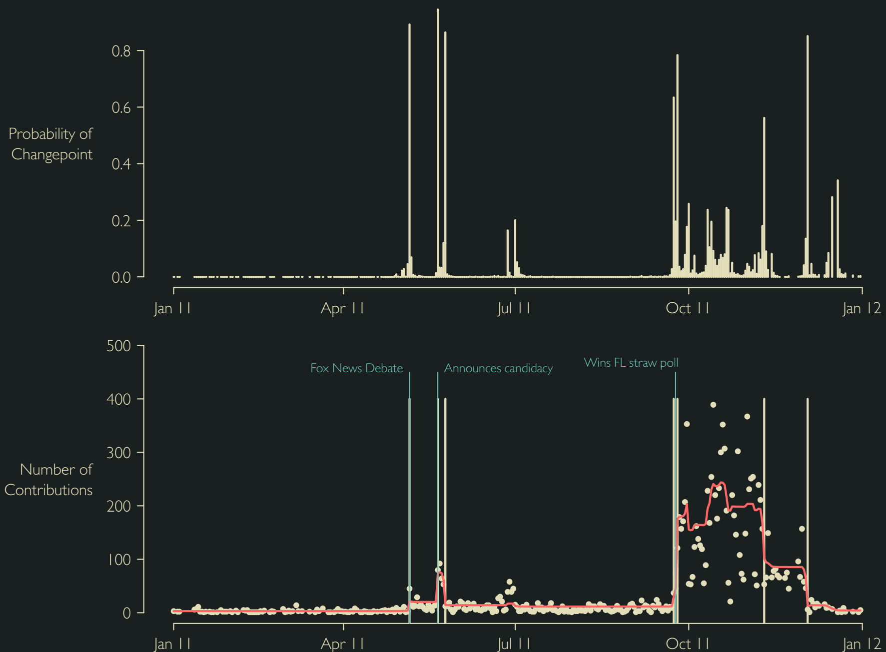


The rise and fall of Herman Cain

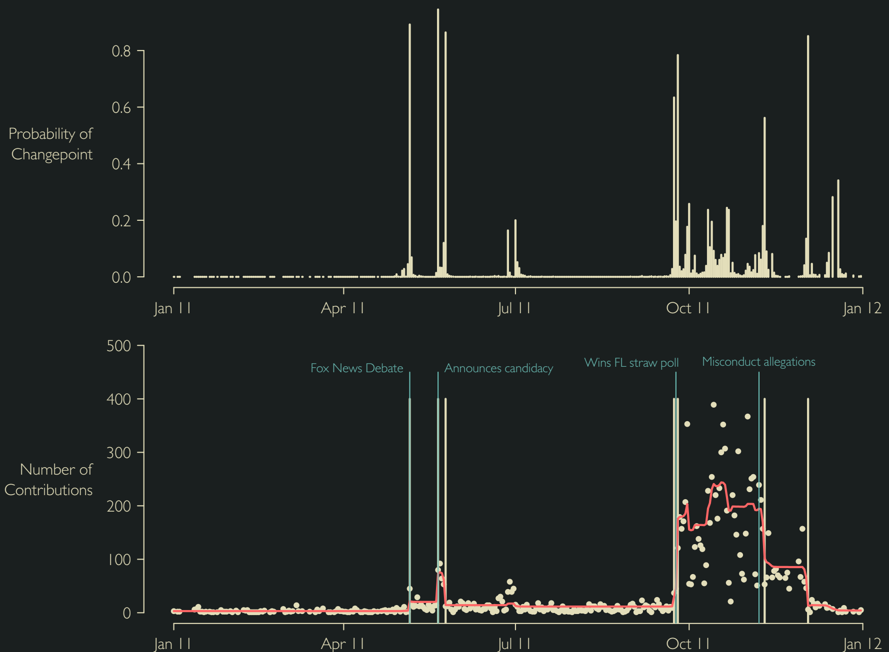
The rise and fall of Herman Cain

The rise and fall of Herman Cain

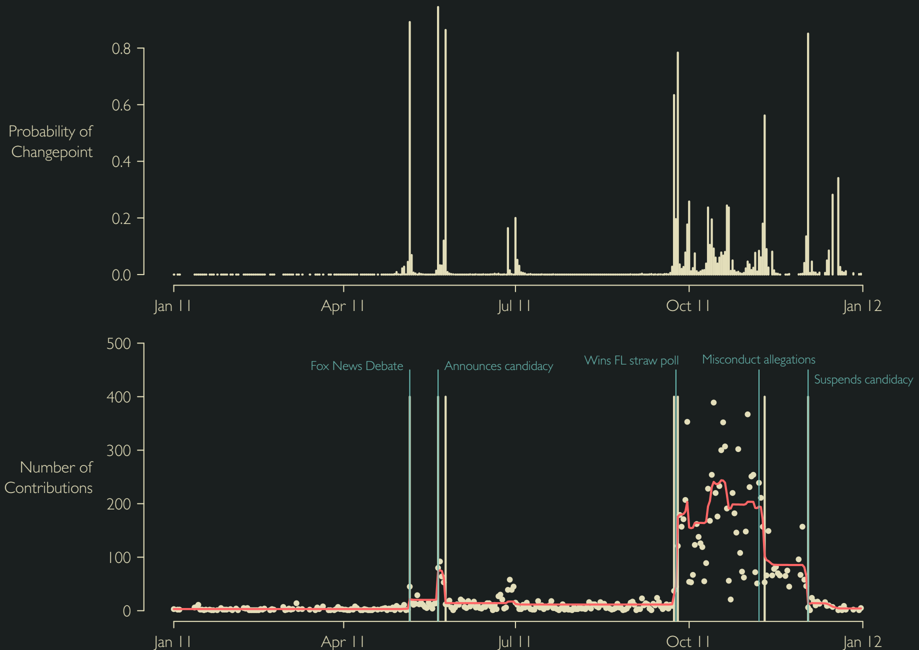
The rise and fall of Herman Cain



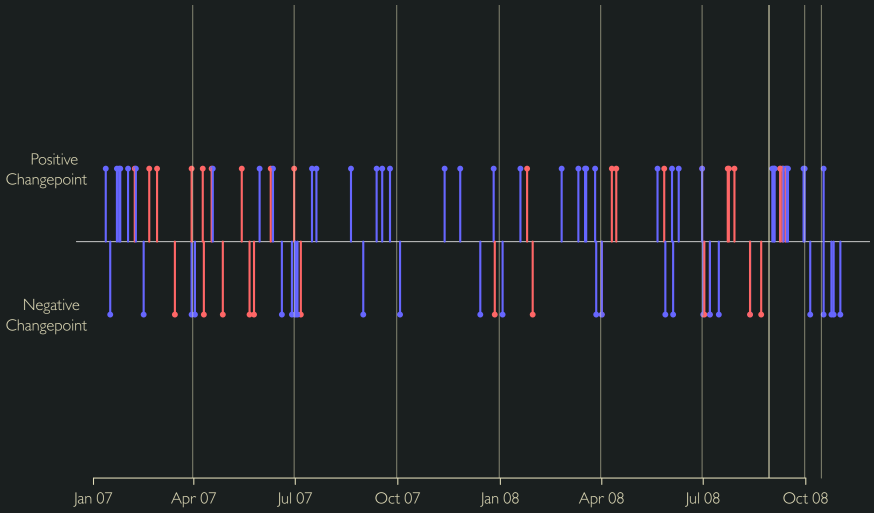
The rise and fall of Herman Cain



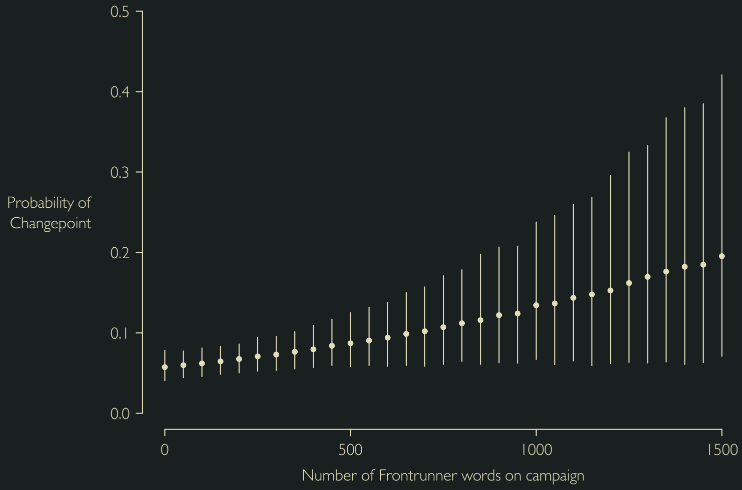
The rise and fall of Herman Cain



All Senate changepoints



More attention around changepoints



The path forward.

The path forward.

1. Run on all (digitzed) Congressional races to find more systematic variation.

The path forward.

1. Run on all (digitzed) Congressional races to find more systematic variation.

2. Compare changepoints for time-series of different types of voters, PACs.

The path forward.

1. Run on all (digitzed) Congressional races to find more systematic variation.

2. Compare changepoints for time-series of different types of voters, PACs.

3. Generalize the Bayesian nonparametric approach beyond count data.