Developments in Positive Empirical Models of Election Frauds: Dimensions and Decisions*

Diogo Ferrari† Kevin McAlister‡ Walter R. Mebane, Jr.§

August 29, 2018

*Prepared for presentation at the 2018 Annual Meeting of the Summer Meeting of the Political Methodology Society, Provo, UT, July 16–18, 2018. Previous versions were presented at the 2016 Summer Meeting of the Political Methodology Society, at the 2016 Annual Meeting of the American Political Science Association at the 2017 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 6–9, 2017, at the 2017 Summer Meeting of the Political Methodology Society, Madison, WI, July 13–15, 2017 and at the 2018 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 5–8, 2018. Work supported by NSF award SES 1523355.

†Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: dferrari@umich.edu).

‡Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: kamcal@umich.edu).

§Professor, Department of Political Science and Department of Statistics, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: wmebane@umich.edu).

Abstract

We present new developments regarding positive empirical models of election frauds. We further develop the general framework for Bayesian estimation of finite mixture models of election frauds (models that include covariates) to support estimating the number of types of frauds. There can be no frauds, one type of fraud or two types. By allowing the dimensionality of the model to vary, fraud components that do not occur completely disappear from the model. We consider how to make statistical decisions about the presence of frauds based on loss functions and posterior predictive distributions of fraud magnitudes. We use the decision framework to compare the varying dimension approach to a fixed-dimensional approach in which the probablities of frauds can become small but always remain positive. Among transdimensional MCMC methods we use reversible jump MCMC but also consider nonparametric Bayes and path sampling methods. We use the model to investigate fraud in several settings, including Kenya and Brazil. Covariates to describe characteristics of voters and polling places are available in the case of Brazil.

1 Introduction

Election forensics is the field devoted to using statistical methods to determine whether the results of an election accurately reflect the intentions of the electors. Election forensics techniques apply statistical methods to low-level aggregates of votes such as polling station or ballot box counts of eligible voters and votes recorded for parties. Many statistical methods for trying to detect election frauds have been proposed (e.g. Myagkov, Ordeshook and Shaikin 2009; Levin, Cohn, Ordeshook and Alvarez 2009; Shikano and Mack 2009; Mebane 2010; Breunig and Goerres 2011; Pericchi and Torres 2011; Cantu and Saiegh 2011; Deckert, Myagkov and Ordeshook 2011; Beber and Scacco 2012; Hicken and Mebane 2015; Montgomery, Olivella, Potter and Crisp 2015). Multimodal distributions of turnout and vote choices are directly or indirectly the basis for several approaches (Myagkov, Ordeshook and Shaikin 2008, 2009; Levin et al. 2009; Rozenas 2017). Mebane (2016) presents a finite mixture likelihood model that implements the sharpest contribution motivated by multimodality, which is a model proposed by Klimek, Yegorov, Hanel and Thurner (2012). All methods for election forensics may be affected by the fundamental ambiguity that it is difficult to distinguish results caused by election frauds from results produced by strategic behavior, but in combination the various methods can estimate, characterize and locate features of low-level aggregates of votes that may indicate frauds (Hicken and Mebane 2015; Mebane 2015).

Klimek et al. (2012) specify functional forms that define two mechanisms by which votes are added to a leading party: manufacturing votes from genuine nonvoters; and stealing votes from the other parties. The "leading party" is the party that the model allows to gain votes from frauds.[1] These mechanisms operate either in an "extreme" manner, so election data have turnout near 100 percent with nearly all votes going to the leader, or in an "incremental" manner, where a substantial number but not almost all votes are reallocated

[1] A limitation of the model is that only one party may be represented as a party that gains votes from frauds.

to the leader. The model includes parameters that express whether vote manufacturing or vote stealing is the predominant form of fraud that occurs. If frauds as described by their model occur, then turnout and vote proportion distributions are bimodal or trimodal.

As Mebane (2016) discusses, the Klimek et al. (2012) idea that multimodalities characterize frauds is motivated by theory developed by Borghesi and Bouchaud (2010). Mebane (2016) argues that the same theory implies that multimodalites are also created by strategic behavior. In both cases the core mechanism is that imitative similarities are induced among electors. Whether frauds and strategic behavior trigger empirical models that focus on multimodality in distinctive ways is a question that Mebane (2016) preliminarily address, but more empirical and theoretical research is needed to understand the distinctions.

The model of Mebane (2016) is based on a conception that closely follows the specification defined by Klimek et al. (2012). We briefly review that specification in section 2.1. The specification uses functional forms that are unusual for political science. As a special case of a general formulation we introduce of the main ideas that motivate Klimek et al. (2012), we describe an alternative specification that may be more familiar to political scientists, in that it includes familiar logistic forms and binomial distributions for turnout and vote choice proportions and counts. In fact Borghesi and Bouchaud (2010) also refer to logits of such proportions, so the alternative model is in that respect more directly in line with their theorizing.

A limitation of this alternative formulation is that it cannot accommodate overdispersion relative to the assumed binomial distributions. Covariates can be incorporated in the model in a straightforward way to affect the distribution of turnout and vote choice means, but overdispersion is still a concern (e.g Wand, Shotts, Sekhon, Mebane, Herron and Brady 2001). Partly for this reason, we also develop a model based on truncated Normal distributions, analogous to the formulation in Mebane (2016).

We adopt a Bayesian approach to formulating the models, which we estimate using

Markov Chain Monte Carlo (MCMC) methods (Plummer 2003; Plummer, Stukalov and Denwood 2016). We present versions that always have all types of frauds present in the model, albeit perhaps only in vestigial amounts, and methods that allow the dimensionality of the model to vary as fraud components appear and disappear as the MCMC proceeds. Reversible jump MCMC (Green 1995) is the approach we use in the current paper. We use the predictive posterior distribution of the magnitude of frauds for each observation to decide—based on explicit loss functions—whether frauds occur, and we assess whether the varying dimension and fixed dimension approaches support making different decisions.

The models allow covariates to condition true voting and turnout behavior as well the patterns of frauds that occur. At least with small datasets, estimation using the model is fast enough that we can perform Monte Carlo simulation exercises in which we estimate the model using simulated data with known characteristics. For one version of the models we report some of these simulation results, including investigation of the frequentist coverage properties of the Bayesian estimates produced using MCMC methods. We also iluustrate using the models with data from elections in Brazil and in Kenya.

2 Models

In the Klimek et al. (2012) formulation the baseline assumption is that votes in an election with no fraud are produced through processes that can be summarized by two Normal distributions: one distribution for turnout proportions and another, independent distribution for the proportion of votes going to the leader. The model conditions on the number of eligible voters. Some votes are transferred to the leader from the opposition, and some are taken from nonvoters. Two kinds of election fraud refer to how many of the opposition and nonvoters votes are shifted: with "incremental fraud" moderate proportions of the votes are shifted; with "extreme fraud" almost all of the votes are shifted. Mebane (2016) and Klimek et al. (2012) have parameters that specify the probability that each unit

experiences each type of election fraud: f_{i} is the probability of incremental fraud and f_{e} is the probability of extreme fraud. Other parameters fully describe bimodal and trimodal distributions that the model characterizes as being consequences of election frauds.

In our general specification the conception of frauds is essentially the same. Instead of using Normal distributions as the fundamental distributions to model turnout and votes for the leader, we have flexibility regarding the distributions that are assumed, and we can specify how the distributions depend on observed covariates. We test a specification that uses binomial distributions and logistic models for those distributions.

2.1 Original Specification

The original Klimek et al. (2012) conception includes three kinds of votes: votes without fraud; votes with "incremental fraud"; and votes with "extreme fraud." With no fraud the distribution of vote counts, given the number of eligible voters, is a truncated product Normal distribution. Fraud corresponds to differing proportions of votes going to the leading party that should have gone to other parties or should not have been counted as votes at all.[2] With incremental fraud a small proportion x_i of nonvotes in electoral unit i are counted for the leading party while a proportion x_i^α, $\alpha > 0$, of votes for opposition instead go to the leading party. With extreme fraud a large proportion $1 - y_i$ of the nonvotes are counted for the leading party and a proportion $(1 - y_i)^\alpha$ of genuine opposition votes instead go to the leading party.

Following are definitions for the original empirical frauds model adapted from Klimek et al. (2012) and implemented in Mebane (2016). Because of its reliance on truncated and folded Normal distributions we refer to this specification as the Normal model.

Observed data come from n electoral units (e.g., polling stations), and the observed number of eligible voters in each unit is N_i, $i = 1, \ldots, n$. Votes for parties are observed as the count of votes for the leading party, denoted W_i, and the sum of votes cast for all other

[2]The "leading" party is not necessarily the party with the most votes. It is the party treated as potentially gaining votes from frauds. Only one party can gain from frauds in the current conception.

parties (the "opposition"), denoted O_i. The number of observed nonvotes ("abstentions") is $A_i = N_i - W_i - O_i$. The observed number of valid votes is $V_i = N_i - A_i$.

Using $\mathcal{N}(\mu, \sigma)$ to denote a normally distributed random variable with mean μ and standard deviation σ, the conceptual specification of the model is as follows. For each electoral unit $i = 1, \ldots, n$, for some $\alpha > 0$ define:

1. true turnout, $\tau_i \sim \mathcal{N}(\tau, \sigma_\tau)$, $0 \le \tau_i \le 1$;

2. the leading candidate's true vote proportion, $\nu_i \sim \mathcal{N}(\nu, \sigma_\nu)$, $0 \le \nu_i \le 1$;

3. no fraud (occurs with probability $f_0 = 1 - f_\mathrm{i} - f_\mathrm{e}$): the number of votes for the leading candidate is $W_i^* = N_i \tau_i \nu_i$ and the number of nonvoters is $A_i^* = N_i (1 - \tau_i)$;

4. incremental fraud (occurs with probability f_i): $x_i \sim |\mathcal{N}(0, \theta)|$, $0 < x_i < 1$, is the proportion of genuine nonvotes counted as votes for the leading candidate and x_i^α is the proportion of votes genuinely cast for others but instead counted as votes for the leading candidate, so the number of votes for the leading candidate is $W_i^* = N_i (\tau_i \nu_i + x_i (1 - \tau_i) + x_i^\alpha (1 - \nu_i) \tau_i)$, and the number of nonvoters is $A_i^* = N_i (1 - x_i) (1 - \tau_i)$;

5. extreme fraud (occurs with probability f_e): using $y_i \sim |\mathcal{N}(0, \sigma_x)|$, $\sigma_x = 0.075$, $0 < y_i < 1$, $1 - y_i$ is the proportion of genuine nonvotes counted as votes for the leading candidate and $(1 - y_i)^\alpha$ is the proportion of votes genuinely cast for others but instead counted as votes for the leading candidate, so the number of votes for the leading candidate is $W_i^* = N_i (\tau_i \nu_i + (1 - y_i) (1 - \tau_i) + (1 - y_i)^\alpha (1 - \nu_i) \tau_i)$, and the number of nonvoters is $A_i^* = N_i y_i (1 - \tau_i)$.

A finite mixture likelihood is defined in Mebane (2016).

2.2 A General Finite Mixture Model of Election Fraud

Let i denote the electoral unit of observation (e.g. a polling station). We continue to use N_i to represent the observed number of eligible voters at i, V_i for the observed count of votes cast, W_i for the observed count of votes for the leader and $A_i = N_i - V_i$ for the number of observed nonvotes ("abstentions"). For each i let τ_i be the true turnout proportion, potentially related to the observed proportion $t_i = V_i/N_i$, and let ν_i be the true proportion of votes for the leader, potentially related to the observed proportion $w_i = W_i/N_i$.

Suppose that either vote stealing or vote manufacturing can happen in an incremental or extreme way. Let ι_i^S and ι_i^M be the proportions of votes incrementally stolen and manufactured, respectively. Define υ_i^S and υ_i^M similarly but for votes stolen or manufactured in an exteme way. Let Z_i be an indicator random variable such that $Z_i = 2$ if incremental fraud has happened in i, $Z_i = 3$ if extreme fraud has happened instead, and $Z_i = 1$ if no fraud has happened in i. The random vector $Y_i = (Z_i, \nu_i, \tau_i, \iota_i^S, \iota_i^M, \upsilon_i^S, \upsilon_i^M)$ is unobservable, and contains the latent variables of the model.

The rules that create the observable proportions w_i and $a_i = 1 - t_i$ are the following:

$$a_i = \begin{cases} 1 - \tau_i & \text{, if } Z_i = 1 \\ (1 - \tau_i)(1 - \iota_i^M) & \text{, if } Z_i = 2 \\ (1 - \tau_i)(1 - \upsilon_i^M) & \text{, if } Z_i = 3 \end{cases} \tag{1}$$

$$w_i = \begin{cases} \tau_i \nu_i & \text{, if } Z_i = 1 \\ \tau_i \nu_i + \iota_i^M (1 - \tau_i) + \iota_i^S \tau_i (1 - \nu_i) & \text{, if } Z_i = 2 \\ \tau_i \nu_i + \upsilon_i^M (1 - \tau_i) + \upsilon_i^S \tau_i (1 - \nu_i) & \text{, if } Z_i = 3 \end{cases} \tag{2}$$

So if no fraud has happened in i, then the observed proportion of turnout and votes for the leader matches the actual values in the election, that is $(w_i, a_i) = (\nu_i \tau_i, 1 - \tau_i)$. Suppose that incremental fraud has happened ($Z_i = 2$). Then the observed abstention is

$a_i = (1 - \tau_i)(1 - \iota_i^M)$, that is, only a fraction $(1 - \iota_i^S)$ of the actual abstention is computed as such, because $\iota_i^S \in [0, 1]$ has been counted as if that proportion of voters had voted to the leader. Additionally, if $Z_i = 2$, the observed votes are not the actual proportions, but the votes after the addition of votes that were stolen from other parties $(\iota_i^S \tau_i (1 - \nu_i))$ and from the abstainers $(\iota_i^M (1 - \tau_i))$. A similar situation occurs when extreme fraud happens, the only difference being the quantities of votes stolen or manufactured.

Note also that w_i can be expressed in terms of a_i instead of τ_i:

$$w_i = \begin{cases} \nu_i (1 - a_i) & , \text{ if } Z_i = 1 \\ \nu_i \left(\dfrac{1 - \iota_i^S}{1 - \iota_i^M} \right) \left(1 - \iota_i^M - a_i \right) + a_i \left(\dfrac{I^M - \iota_i^S}{1 - \iota_i^M} \right) + \iota_i^S & , \text{ if } Z_i = 2 \\ \nu_i \left(\dfrac{1 - \upsilon_i^S}{1 - \upsilon_i^M} \right) \left(1 - \upsilon_i^M - a_i \right) + a_i \left(\dfrac{\upsilon_i^M - \upsilon_i^S}{1 - \upsilon_i^M} \right) + \upsilon_i^S & , \text{ if } Z_i = 3 \end{cases} \tag{3}$$

These transformations capture the types of frauds discussed in the previous section. As can be seen, there are no distributional assumptions except that the distributions must reflect the ranges of the random variables.

We can provide a parametric model for the unobserved variables in the following way. Let $\mathcal{P} = \left\{ f_{\mu\sigma} : f_{\mu\sigma} = f(y \mid \mu, \sigma), \mathbb{E}[Y] = \mu, \mathbb{V}\text{ar}[Y] = \sigma, \text{ for } \mu, \sigma \in \Omega \right\}$ denote a parametric family of density functions indexed by the expectation and the variance of the random variables whose behavior they describe. So we denote $Y \sim f(y \mid \mu_y, \sigma_y)$. We let the expectations and variances be functions of characteristics $X_i = (X_{0i}, ..., X_{(d-1)i})$ of the electoral unit i. The random vector X_i encodes characteristics of the electorate that vote in i and of the electoral unit i itself. We impose the assumption that the unobseved variabels are independent and indentically distributed given that we condition them on the characteristics of the voters and the electoral units. That is, $\forall i \neq j, (Y_i \perp\!\!\!\perp Y_j) \mid X_i$. Denote

the distribution of the unobserved random variables by:

$$
\begin{aligned}
Z_i \mid X_i &\sim \mathrm{Cat}(\boldsymbol{\pi})\,, \quad Z_i \in \{1,2,3\} \\
\tau_i \mid X_i &\sim f(\tilde{t}_i \mid \mu_\tau, \sigma_\tau)\,, \quad \tau_i \in [0,1] \\
\nu_i \mid X_i &\sim f(\tilde{w}_i \mid \mu_\nu, \sigma_\nu)\,, \quad \nu_i \in [0,1] \\
\iota_i^l \mid X_i &\sim f(i_i^l \mid \mu_\iota^l, \sigma_\iota^l)\,, \quad \iota_i^l \in [0,k_1]\,, l \in \left\{S, M\right\}, k_1 \in [0,1] \\
\upsilon_i^l \mid X_i &\sim f(e_i^l \mid \mu_\upsilon^l, \sigma_\upsilon^l)\,, \quad \upsilon_i^l \in [k_2,1]\,, k_2 \in [0,1]\,, k_1 \leq k_2
\end{aligned}
\tag{4}
$$

where for $y \in \{\tau, \nu, \iota, \upsilon\}$,

$$
\begin{aligned}
\mu_y &= h_y^{(1)}(X_i) \\
\sigma_y &= h_y^{(2)}(X_i)\,.
\end{aligned}
$$

Denote $\boldsymbol{\theta} = (\pi, \boldsymbol{\mu}, \boldsymbol{\sigma})$, $\boldsymbol{\mu} = (\mu_\nu, \mu_\tau, \mu_\iota^l, \mu_\upsilon^l)$, $\boldsymbol{\sigma} = (\sigma_\tau, \sigma_\nu, \sigma_\iota^l, \sigma_\upsilon^l)$. We represent the proportions (w_i, a_i) as taking values in $C = [0,1]^2$. For a choice of specific distributions $f_{\mu\sigma} \in \mathcal{P}$, we have that the sampling distribution of a single pair (w_i, a_i) is given by:

$$
f(w_i, a_i \mid \boldsymbol{\theta}) = \begin{cases}
\displaystyle\prod_{j \in \{w_i, t_i\}} \frac{f(g_j^{-1} \mid \mu_j, \sigma_j)}{(1-a_i)}\,, & \text{if } Z_i = 1 \\
\displaystyle\int_C \prod_{j \in \{w_i, t_i, \iota_i^S, \iota_i^M\}} \frac{f(g_j^{-1} \mid \mu_j, \sigma_j)}{(1-\iota_i^M - a_i)(1-\iota_i^S)}\, m(d\iota^S \times d\iota^M)\,, & \text{if } Z_i = 2 \\
\displaystyle\int_C \prod_{j \in \{w_i, t_i, \upsilon_i^S, \upsilon_i^M\}} \frac{f(g_j^{-1} \mid \mu_j, \sigma_j)}{(1-\upsilon_i^M - a_i)(1-\upsilon_i^S)}\, m(d\upsilon^S \times d\upsilon^M)\,, & \text{if } Z_i = 3
\end{cases}
\tag{5}
$$

where $m(d\iota^S \times d\iota^M)$ and $m(d\upsilon^S \times d\upsilon^M)$ are product measures and

$$
\begin{aligned}
g^{-1}_{\iota_i^S} &= i_i^S \\
g^{-1}_{\iota_i^S} &= i_i^M \\
g^{-1}_{t_i} = \tau_i &= \begin{cases} 1 - a_i & \text{if } Z = 1 \\ 1 - \dfrac{a_i}{1 - \iota_i^M} & \text{if } Z = 2 \\ 1 - \dfrac{a_i}{1 - \upsilon_i^M} & \text{if } Z = 3 \end{cases} \\
g^{-1}_{w_i} = \nu_i &= \begin{cases} \dfrac{w_i}{1 - a_i}\,, & \text{if } Z = 1 \\ w_i \left(\dfrac{1}{1 - \iota_i^M - a_i} \right) \left(\dfrac{1 - \iota_i^M}{1 - \iota_i^S} \right) - \dfrac{\iota_i^S}{1 - \iota_i^S} - \dfrac{a_i \iota_i^M}{(1 - \iota_i^M - a_i)(1 - \iota_i^S)}\,, & \text{if } Z = 2 \\ w_i \left(\dfrac{1}{1 - \upsilon_i^M - a_i} \right) \left(\dfrac{1 - \upsilon_i^M}{1 - \upsilon_i^S} \right) - \dfrac{\upsilon_i^S}{1 - \upsilon_i^S} - \dfrac{a_i \upsilon_i^M}{(1 - \upsilon_i^M - a_i)(1 - \upsilon_i^S)}\,, & \text{if } Z = 3 \end{cases}
\end{aligned}
$$

Details of the derivation can be found in the appendix. The model just described represents a generalization of the finite mixture model underlying the arguments put forward by Klimek et al. (2012). In the following sections, we develop and implement two models based on different choices of $f_{\mu\sigma} \in \mathcal{P}$ and some additional restrictions on the random variables and parameters.

2.2.1 The Logistic-Binomial Finite Mixture Model

One parametrization uses a binomial model for $f_{\mu\sigma}$. The distributions then are given by

$$
\begin{aligned}
Z_i &\sim \text{Cat}(\boldsymbol{\pi}), \qquad \boldsymbol{\pi} = (\pi_0, \pi_I, \pi_E) \in \Delta^3 \\
N_i \nu_i &\sim Bin(\nu, N_i) \\
N_i \tau_i &\sim Bin(\tau, N_i) \\
N_i \iota_i^l &\sim Bin(\mu_\iota^l, N_i), \quad \zeta_i < k \in [0,1], \quad l \in \{S, M\} \\
N_i \upsilon_i^l &\sim Bin(\mu_\upsilon^l, N_i), \quad \chi_i \geq k \in [0,1],
\end{aligned}
\tag{6}
$$

The parameters μ_ι^l and μ_ι^l are fraud parameters at the unit level observation. They capture, respectively, the probability of incremental fraud and extreme fraud at the unit i.

The parameter ν captures the expected proportion of votes for the leader candidate at i, and τ represents the expected turnout proportion in i. Same for μ_ι^l and μ_υ^l. However, these parameters now are in the unit interval. So we defined them as

$$\tau = \frac{1}{1+e^{-\gamma^T x_i}} \quad , \quad \gamma \in \mathbb{R}^d \quad ; \quad \mu_\nu = \frac{1}{1+e^{-\beta^T x_i}} \quad , \quad \beta \in \mathbb{R}^d \tag{7}$$

$$\mu_\iota^l = \frac{1}{1+e^{-\rho_l^T x_i}} \quad , \quad \rho_l \in \mathbb{R}^d \quad ; \quad \mu_i^l = \frac{1}{1+e^{-\delta_l^T x_i}} \quad , \quad \delta_l \in \mathbb{R}^d \tag{8}$$

One choice for the hyperprior on those parameters is:

$$\rho_l \sim N_d(\mu_\rho, \xi I_{d\times d}) \qquad \beta \sim N_d(\mu_\beta, \xi I_{d\times d})$$

$$\delta_l \sim N_d(\mu_\delta, \xi I_{d\times d}) \qquad \gamma \sim N_d(\mu_\gamma, \xi I_{d\times d}), \quad \xi \in \mathbb{R}_+$$

$$\pi \sim \text{Dirichlet}(\psi_1, \psi_2, \psi_3), \quad \psi_1, \psi_2, \psi_3 \in R_+$$

The assumptions on the dependence structure of the random variables in the model are, $\forall i, j = 1, ..., N$ and $i \neq j$, $(A1)(\nu_i \perp\!\!\!\perp \nu_j) \mid X$, $(A2)(A_i \perp\!\!\!\perp A_j) \mid X$, $(A3)(\nu_i \perp\!\!\!\perp A_j) \mid X$, $(A4)\iota_i^l \perp\!\!\!\perp \upsilon_i^l \perp\!\!\!\perp Z_i$ The assumptions (A1) to (A3) mean that, after taking into account the covariates and their effect on the expected turnout and on the expected support for a specific party in i, these two elements on an unit j have not effect on the ones in i. (A4) says that incremental and extreme fraud in i are independent of the expected fraud in the election. In other words, it says that if there is a positive probability of incremental fraud in the election, that probability does not affect the chance of fraud in a specific unit i in any way more than it does in j.

2.2.2 Restricted Normal Finite Mixture Approximation

In this section, we impose restrictions on the general model presented above to come close to the models used by Klimek et al. (2012) and Mebane (2016).[3] The current Bayesian model expresses a genuine joint distribution for turnout and votes for the leader, unlike the model defined in Mebane (2016). We let the turnout and vote choice mean parameters be linear functions of covariates.

Contrary to the binomial model, this model assumes that the probability of incremental fraud is linked to the probability of extreme fraud: the two kinds of frauds are connected by the α parameter of the model.

Define, for $\alpha > 0$, $\iota = \iota^M$, $\iota^\alpha = \iota^S = (\iota^M)^\alpha$, $\upsilon = \upsilon^M$, and $\upsilon^\alpha = \upsilon^S = (\upsilon^M)^\alpha$. That is, the proportion of votes stolen, in an incremental or extreme fashion, and the proportions of votes manufactured are related by a positive parameter α. Let $f_{\mu\sigma}$ be a restricted normal distribution with support $[0, 1] \subset \mathbb{R}$. Denote $\phi(y \mid \mu_y, \sigma_y)$ and $\Phi(y \mid \mu_y, \sigma_y)$ the gaussian density and the gaussian measure, respectively, and $\phi(y \mid \mu_{\mathrm{y}} \, , \, \sigma_{\mathrm{y}} \, , \, 0, \, 1)$ and $\Phi(y \mid \mu_{\mathrm{y}} \, , \, \sigma_{\mathrm{y}} \, , \, 0, \, 1)$ their restricted version on the unit interval of the real numbers. We let $\mu_\tau = X_i^T \boldsymbol{\gamma}$ and $\mu_\nu = X_i^T \boldsymbol{\beta}$, that is, the parameters of the unrestricted normal distribution of the proportion of votes for the leader and the turnout are functions of covariates. Denote $\boldsymbol{\theta}' = (\boldsymbol{\theta}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$. In order to define the Bayesian version of the model, we treat $(w_i, a_i)_{i=1}^n$ as transformations of $\{\nu_i, A_i\}_{n=1}^n$ exclusively as given by equations (1) and (3). We condition its distribution on all the other quantities. By doing this, we avoid the necessity to compute integrals over ι_i and υ_i in the sampling distribution as expressed by the equations in (5), as was necessary in Mebane (2016). By conditioning the sample distribution on ι_i and υ_i, such integration is naturally approximated by the MCMC procedure. Now we are ready to define the sampling distribution of $(w_i, a_i)_{i=1}^n$ for the

[3] To represent frauds Klimek et al. (2012) and Mebane (2016) use a folded and truncated normal distribution, $x_i \sim |\mathcal{N}(0, \theta)|$, while the restricted Normal Bayesian formulation uses a simply truncated Normal distribution, $\iota_i^l \mid X_i \sim f(i_i^l \mid \mu_\iota^l, \sigma_\iota^l)$, $i_i^l \in [0, 1]$.

Bayesian Restricted Normal Finite Mixture (BRNFM) model:

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{a} \mid Z_i, \iota_i, \upsilon_i, \boldsymbol{\theta'}) = \prod_{i=1}^{n} f(w_i \mid a_i, Z_i, \iota_i, \upsilon_i, \boldsymbol{\theta'}) f(a_i \mid Z_i, \iota_i, \upsilon_i, \boldsymbol{\theta'}) \tag{9}$$

where

$$f(a_i \mid Z_i, \iota_i, \upsilon_i, \boldsymbol{\theta'}) \begin{cases} \phi\left(g_{Ti}^{-1} \mid \mu_\tau, \sigma_\tau, 0, 1\right) & , \text{ if } Z_i = 1 \\ \phi\left(g_{Ti}^{-1} \mid \mu_\tau, \sigma_\tau, 0, 1\right)\left(\dfrac{1}{1-\iota_i}\right) & , \text{ if } Z_i = 2 \\ \phi\left(g_{Ti}^{-1} \mid \mu_\tau, \sigma_\tau, 0, 1\right)\left(\dfrac{1}{1-\upsilon_i}\right) & , \text{ if } Z_i = 3 \end{cases} \tag{10}$$

and

$$f(w_i \mid a_i, Z_i, \iota_i, \upsilon_i, \boldsymbol{\theta'}) = \begin{cases} \phi\left(g_{wi}^{-1} \mid \mu_\nu, \sigma_\nu, 0, 1\right)\left(\dfrac{1}{1-a_i}\right) & , \text{ if } Z_i = 1 \\ \phi\left(g_{wi}^{-1} \mid \mu_\nu, \sigma_\nu, 0, 1\right)\left(\dfrac{1-\iota_i}{(1-\iota_i-a_i)(1-\iota_i^\alpha)}\right) & , \text{ if } Z_i = 2 \\ \phi\left(g_{wi}^{-1} \mid \mu_\nu, \sigma_\nu, 0, 1\right)\left(\dfrac{1-\upsilon_i}{(1-\upsilon_i-a_i)(1-\upsilon_i^\alpha)}\right) & , \text{ if } Z_i = 3 \end{cases} \tag{11}$$

The variables transformations g^{-1} are given as before, with the modifications to $\iota^M, \iota^S, \upsilon^S$ and υ^M introduced in this section. This is a straghtforward result that follows from the joint density in (5). Details are in the appendix. The prior distributions of the latent variables Z_i, υ_i, and ι_i are as in (4) above, with f in each case being a restricted normal distribution. Additionally, we use the following disperse distributions for the

hyperparameters:

$$
\begin{aligned}
\pi \sim & Dirichlet(\psi) \quad , \\
\alpha \sim & \Gamma(0.01, 0.01) \\
\boldsymbol{\beta} \sim & \phi_d(\beta \mid \mu_\beta, \sigma_\beta) \\
\boldsymbol{\gamma} \sim & \phi_d(\gamma \mid \mu_\gamma, \sigma_\gamma) \\
\xi_l \sim & \Gamma(0.01, 0.01) \quad , \quad l \in \{w, T, I\} \quad , \quad \xi_l = \frac{1}{\sigma_l^2}
\end{aligned}
\tag{12}
$$

We define the following constants:

$$
\begin{aligned}
\mu_I = 0 \quad , \quad \mu_\upsilon = 1 \quad , \quad \sigma_\upsilon = 0.075 \quad , \quad \psi = (1, 1, 1) \\
\mu_\beta = 0 \in \mathbb{R}^d \quad , \quad \sigma_\beta = \mathbb{I}_{d \times d} \quad , \quad \mu_\gamma = 0 \in \mathbb{R}^d \quad , \quad \sigma_\gamma = \mathbb{I}_{d \times d}
\end{aligned}
$$

As mentioned, we also have

$$
\nu = X_i^T \beta \quad , \quad \mu_\tau = X_i^T \gamma
$$

2.3 MCMC Details

We use two stopping rules for the MCMC to improve reproducibility of the results and as a minimal convergence diagnostic requirement. The first is the Potential Scale Reduction Factor (PSRF), which was required to be at most 1.1 (Brooks, Gelman, Jones and Meng 2011; Gelman and Shirley 2011; Gelman and Rubin 1992).

The second is the univariate Markov Chain Monte Carlo Standard Error (MCMCSE) (Flegal, Haran and Jones 2008; Flegal and Hughes 2012; Gong and Flegal 2015; Vats, Flegal and Jones 2015). For a posterior sample $\{\theta_i\}_{i=1}^n$ and the posterior expectation denoted by $\mathbb{E}[\boldsymbol{\theta} \mid \cdot]$, under mild regularity conditions, $(1/n)\sum_{i=1}^n \boldsymbol{\theta}_i = \overline{\boldsymbol{\theta}} \xrightarrow{p} \mathbb{E}[\boldsymbol{\theta} \mid \cdot]$ (Meyn and Tweedie 2012; Robert and Casella 2013). Additionally, when the Markov Chain

satisfies some ergodicity requirements, then $\sqrt{n}(\overline{\boldsymbol{\theta}} - \mathbb{E}[\boldsymbol{\theta} \mid \cdot]) \xrightarrow{d} N(\mathbf{0}, \Sigma)$, where $[\Sigma]_{kk} = \sigma_{\theta_k} = \sqrt{\mathbb{V}\text{ar}[\theta_{k1}] + 2\sum_{i=2}^{\infty} \mathbb{C}\text{ov}[\theta_{k1}, \theta_{ki}]}$ is the asymptotic variance of the k^{th} component of $\boldsymbol{\theta}$. We implement the MCMC using a Metropolis-Hasting algorithm, which is known to satisfy such ergodicity requirements. So we can run the chain long enough to make the asymptotic variance of the MCMCSE—i.e. variance of $\overline{\boldsymbol{\theta}} - \mathbb{E}[\boldsymbol{\theta} \mid \cdot]$—as small as we want (Meyn and Tweedie 2012; Flegal, Haran and Jones 2008; Flegal 2008). We ran the chains until the asymptotic variance of the MCMCSE was smaller than 0.02. We estimate σ_{θ_k} using consistent nonoverlaping batch means (for details see Jones, Haran, Caffo and Neath (2006); Flegal, Haran and Jones (2008)), which provides a consistent estimator for σ_{θ_k}. We use univariate MCMCSE, which is more conservative than the multivariate case (Vats, Flegal and Jones 2015).

Other usual diagnostics are also performed when there is sign of bad mixing (Cowles and Carlin 1996).

By the same convergence in distribution argument just outlined, as $n \to \infty$, $(1/n)\sum_{i=1}^{n} \overline{\boldsymbol{\theta}}_i \xrightarrow{p} \mathbb{E}[\boldsymbol{\theta} \mid \cdot] \xrightarrow{p} \boldsymbol{\theta}_e$, where $\boldsymbol{\theta}_e$ is the (fixed) value that generated the data (Gelman, Carlin, Stern and Rubin 2014, 111). So we conduct a frequentist assessment of performance of the Bayesian procedure by computing a coverage probability defined as $\text{Cov}(\boldsymbol{\theta}_e) = p_{(\nu,\tau)|\boldsymbol{\theta}_e}(\boldsymbol{\theta}_e \in \mathcal{I}(\hat{\nu}, \hat{\tau}))$ for $\mathcal{I}(\hat{\nu}, \hat{\tau}) = (\boldsymbol{\theta}_L(\hat{\nu}, \hat{\tau}), \boldsymbol{\theta}_U(\hat{\nu}, \hat{\tau}))$ (Carlin and Louis 2009). We construct three intervals $\mathcal{I}(\hat{\nu}, \hat{\tau})$, one using a 95% Highest Posteiror Density, the 95% interval using the asymptotic normal approximation for $\overline{\boldsymbol{\theta}}$, and the empirical confidence interval using order statistics. See Bayarri and Berger (2004) for a discussion and comparison of these methods to construct confidence intervals and coverage probabilities.

2.3.1 Identifying Clusters via Reversible Jump MCMC

In this section, we discuss and derive methods for determining which clusters are needed to model a set of vote data. Each of the previous models utilizes clustering approaches to place observed vote counts into groups representing no fraud, incremental fraud, and

extreme fraud (Klimek et al. 2012; Mebane 2016). Each observation maps to a vector that dictates the probability that it belongs to each possible cluster and these vectors are related to the global mixing vector, π, that can be interpreted as the average probability that each observation, *a priori*, comes from each cluster. π is assumed to follow a Dirichlet distribution and is modeled using the standard mixture modeling approach.

An important question to be asked of vote data is the probability that **any** fraud occurs. In other words, it is important to examine the probability that any observations belong to the incremental or extreme fraud clusters. Under standard mixture model specifications, this question is impossible to answer explicitly—because the global mixing parameters are Dirichlet distributed, $P(\pi_j = 0) = 0$. Thus, saying that a cluster is not needed uses a heuristic that values of π_j near zero indicate that the cluster is not needed. However, this can also be explained by a small number of observations, relative to the number of total observations, having a high probability of belonging to a fraudulent group. In order to effectively determine if any fraud exists, there needs to be a more rigorous examination of these probabilities.

Effectively separating the two cases of zero requires further refinement of the global mixing probabilities. Formally, define the global mixing parameters:

$$P(\pi_j) \sim z_j \; Gamma\left(\frac{\psi_j}{\sum_j \psi_j}, 1\right) + (1 - z_j)\delta_0 \tag{13}$$

$$P(z_j) \sim Bernoulli(\pi_{0,j}) \tag{14}$$

where ψ_j is the prior parameter of the Dirichlet distribution associated with cluster j, z_j is a binary random variable, δ_0 a Dirac delta at 0, and $\pi_{0,j}$ is a prior hyperparameter associated with Bernoulli random variable [4]. This construction is similar to that of the Beta-Bernoulli process defined by Hjort (1990). In short, this specification induces a *spike*

[4]Without loss, the Dirichlet distribution can be represented as a function of Gamma distributions. This allows for easier manipulation of the individual parameters within the Dirichlet distribution.

and slab prior on the elements of the Dirichlet distribution: z_j takes on a value of one and the parameter is freely modeled if and only if the value is far enough away from zero; otherwise, it is fixed at zero. When conditioned on data, the magnitude of the value and the error distribution associated with the value are taken into account when determining the value of the binary variable. Note that this construction explicitly addresses the zero problem posed above by separating the idea of small values and "true" zeros.

Estimating z_j is a challenging exercise. Conjugacy of the beta and Bernoulli priors allows for simple sampling schemes, but the theory behind the sampling schemes is challenging and can lead to difficulty uncovering efficient, unbiased estimates. To this end, three approaches are examined: a grid based approach using path sampling to approximate a Bayes factor; Bayesian nonparametric approaches; and transdimensional MCMC approaches.

A grid based approach is an attractive option since there are a limited number of potential models. Utilizing the path sampling scheme shown by Gelman and Meng (1998), each of the four potential models can be examined and the Bayes factors comparing each can be determined [5] Though this approach is relatively simple and provides a succinct summary of which model is the best given the data, there are two main weaknesses. First, the model requires a complex derivative over the likelihood function. Though the bridge between models is simple to construct, taking the derivative in the high dimensional parameter space is a challenge, particularly for the restricted normal formulation. Second, this approach is incredibly demanding computationally. As the desired accuracy of the Bayes factor increases, the number of points from each bridge to be examined must also increase. Given the potentially multimodal nature of the resulting posteriors, convergence across chains can be time consuming. Multiplying this time by number of chains, then number of bridge points, then across all needed comparisons leads to an incredibly

[5] In these models, we always assume that the "no fraud" cluster is present. Thus, model comparison methods will only consider four viable models: no fraud, no fraud + incremental fraud, no fraud + extreme fraud, and the full model.

demanding computational routine. Though computing resources are abundant, this routine requires computing power that is not readily available. For these reasons, we consider other approaches.

A second potential approach utilizes Bayesian nonparametric solutions to modeling z_j. On this front, there are two potential process priors that can be used: a beta process approach (Paisley and Carin 2009; Paisley, Zaas, Woods, Ginsburg and Carin 2010; Paisley, Blei and Jordan 2012; Paisley and Jordan 2016) and a Dirichlet process approach (Ferguson 1973; Blackwell and MacQueen 1973; Rasmussen 2000; Pitman et al. 2002). Each of these approaches have the same flavor: an infinite number of potential clusters can exist and only a sparse set are active at any given point in time. Both of these approaches provide a nuanced way to model the number of clusters as part of the MCMC rather than relying on post-processing techniques. This approach is promising, especially with work on stick breaking processes that lead to simpler MCMC solutions (Ishwaran and James 2001; Dunson and Park 2008; Teh 2011; Gershman and Blei 2012). These approaches provide a computationally efficient way to determine the number of clusters needed to model vote return data. However, this approach relies on the potential existence of a large number of clusters to work. In our case, there are only three potential clusters with specific known formulations. While there has been some work in the area of truncated Dirichlet processes, the results are known to yield potentially biased estimates of the mixing parameters. This finding was corroborated by our own simulations. Though this problem tends to disappear when $N \to \infty$, unbiased results in simulations are preferred to computational efficiency and another approach was utilized. Further work in this budding area may lead to usage in the future.[6]

The approach we chose utilizes reversible jump MCMC (RJMCMC) to model the number of clusters within the MCMC routine. RJMCMC (Green 1995) assumes that there are M potential models that are to be explored by the MCMC routine. Unlike model

[6]Indeed, we use a beta-binomial approach when illustrating how to make statistical decisions about whether frauds occur, in section 3.1.4.

comparison methods that utilize Bayes factors, RJMCMC models the transition matrix associated with the set of models and models the probabilities that each model is the best, given the data, in conjunction with the model parameters. Given M potentially nested models to be tested, the challenge in creating a RJMCMC routine is determining how to make the decision to switch from one model to another. In our particular case, we propose a routine that first checks if the incremental fraud cluster is needed and, then, if the extreme fraud cluster is needed. These decisions are made sequentially and conditional on the previous state.

Formally, each decision is made using a Metropolis-Hastings step. In the case where the current model includes one of the clusters, the proposed model for comparison sets each of the parameters related to the cluster of interest to zero. When the current model does not include a cluster, the values of parameters related to the cluster are drawn from proposal distributions, $\xi(\cdot)$. To this end, terminal nodes are drawn from conjugate uniform or diffuse distributions and any other values are drawn from their conditional distributions. Let Θ be the parameter values from the current state of the model and Θ' be the proposed parameter values. Then, each MH step requires drawing from:

$$P(\Theta') = \min \left\{ \frac{\prod_{i=1}^{N} P(w_i|a_i, \Theta')P(a_i|\Theta')\xi(\Theta)}{\prod_{i=1}^{N} P(w_i|a_i, \Theta)P(a_i|\Theta)\xi(\Theta')}, 1 \right\} \tag{15}$$

This is the typical MH draw, where the alternative is always accepted if it is better than the previous state or accepted with some probability if it is a worse fit. The choice of uniform proposal distribution can be improved, but performance seems to be minimally impacted: rate of convergence seems to be the aspect most affected by this decision.

2.3.2 RJMCMC Computation

Deriving the sampling schemes necessary to draw from posteriors that utilize transdimensional MCMC steps can be a challenging exercise. In particular, the sampling strategy used to move between models in RJMCMC can have major implications on the overall convergence of the procedure. One approach that ensures correct sampling is to use canned MCMC software, like `JAGS` (Plummer 2010). However, `JAGS` is not flexible enough to accommodate more complex sampling schemes like those being explored in this paper. For this reason, we utilize `NIMBLE` (de Valpine, P., Turek, Paciorek, Anderson-Bergman, Lang and Bodik 2017), a new C++ wrapper package for use in `R`.

Like `JAGS`, `NIMBLE` provides automated sampling procedures generated from an undirected, acyclic graph specified by the user. Unlike `JAGS`, `NIMBLE` allows the user to specify functions in C++ that can be subsequently used with the sampling procedures built into the standard `NIMBLE` library. This provides two major advantages. First, the C++-alike nature of the sampler allows the sampling procedure to be run from compiled code. This provides a significant speed increase over similar procedures in the `JAGS` language. Second, `NIMBLE` provides a method for fine tuning the often costly RJMCMC steps within the sampler—rather than using the default slice sampler at each RJMCMC step, `NIMBLE` allows custom samplers to be created and used within the sampler. As with the compiled code, this also provides speed increases in terms of CPU time and decreased convergence time. All in all, `NIMBLE` provides a sampling approach that is flexible and fast and allows more complicated sampling schemes to run in a reasonable amount of time, even with larger sample sizes.

2.4 Deciding Whether Frauds Occur

A decision whether frauds occur, based on these models, should be informed by the stakes involved in the decision. Deciding that frauds occur implies that some kinds of remedies should be implemented and that some kinds of sanctions should be levied, presumably

against perpetrators and beneficiaries. Deciding there are no frauds implies the election should be celebrated or at least legitimated. The consequences of deciding wrongly can entail misapplied penalties or wrongly seated officials.

Bayesian decision theory offers a framework to combine estimation uncertainties with specified costs and benefits when deciding whether frauds occur. The simplest approach is to use what Berger (1985, 63–64) describes as "$0 - K_i$" loss functions. We imagine there are two actions, a_C and a_F, where a_C is the action taken when there is a decision that there are no frauds and a_F is the action taken when there is a decision that frauds occur. We define two sets of parameter values, Θ_F and Θ_C, where the values in Θ_F imply there are frauds and the values in Θ_C imply there are no frauds. We consider Θ_F and Θ_C to be numbers or proportions of fraudulent votes, according to the model. In the strictest case $\Theta_C = 0$, but generally small numbers or fractions of model-fraudulent votes will be considered not big enough to trigger a_F so Θ_C will be an interval. For $K_C > 0$ and $K_F > 0$ the "$0 - K_i$" loss function has the form

$$L(\theta, a_C) = \begin{cases} 0 & \text{if } \theta \in \Theta_C, \\ K_C & \text{if } \theta \in \Theta_F, \end{cases} \tag{16a}$$

$$L(\theta, a_F) = \begin{cases} 0 & \text{if } \theta \in \Theta_F, \\ K_F & \text{if } \theta \in \Theta_C \end{cases} \tag{16b}$$

where the losses are normalized so that the loss is positive when the decision is wrong.

The idea is to choose the action that has the smallest posterior expected loss, where the probability distributions used to evaluate the posterior losses come from the estimation. Using x to denote the observed data, the posterior expected losses are (Berger 1985,

163–164)

$$E^{\pi(\theta|x)}[L(\theta, a_C)] = K_C \int_{\Theta_F} dF^{\pi(\theta|x)}(\theta) = K_C P(\Theta_F \mid x) \tag{17a}$$

$$E^{\pi(\theta|x)}[L(\theta, a_F)] = K_F \int_{\Theta_C} dF^{\pi(\theta|x)}(\theta) = K_F P(\Theta_C \mid x)\,. \tag{17b}$$

So decide there are frauds if and only if $K_F P(\Theta_C \mid x) < K_C P(\Theta_F \mid x)$.

Decisions that frauds occur are most consequential when made about individual polling stations or about whatever other aggregation unit supplies the observations used in the analysis. It is for such units that enacting remedies or imposing sanctions is likely to be realistically feasible. For each aggregation unit j we consider losses $L(\theta_j, a_{Cj})$ and $L(\theta_j, a_{Fj})$ with decisions based on $K_{Fj} P(\Theta_{Cj} \mid x)$ and $K_{Cj} P(\Theta_{Fj} \mid x)$.

It is possible also to consider decisions that calibrate losses that vary based on the magnitude of frauds. Then for some functions $K_C(\theta)$ and $K_F(\theta)$ where $\partial K_C(\theta)/\partial\theta > 0$ and $\partial K_F(\theta)/\partial\theta > 0$ loss functions may take the form (Berger 1985, 63-64)

$$L(\theta, a_C) = \begin{cases} 0 & \text{if } \theta \in \Theta_C, \\ K_C(\theta) & \text{if } \theta \in \Theta_F, \end{cases} \tag{18a}$$

$$L(\theta, a_F) = \begin{cases} 0 & \text{if } \theta \in \Theta_F, \\ K_F(\theta) & \text{if } \theta \in \Theta_C \end{cases} \tag{18b}$$

and the posterior expected losses are

$$E^{\pi(\theta|x)}[L(\theta, a_C)] = \int_{\Theta_F} K_C(\theta) dF^{\pi(\theta|x)}(\theta) \tag{19a}$$

$$E^{\pi(\theta|x)}[L(\theta, a_F)] = \int_{\Theta_C} K_F(\theta) dF^{\pi(\theta|x)}(\theta)\,. \tag{19b}$$

In any case we use the MCMC chains to evaluate the posterior expectations.

3 Analysis of Simulated Data

This section uses simulated data sets to evaluate the performance of the logistic-binomial model. We conduct a Monte Carlo exercise using samples from a fixed-dimensional Markov Chain to evaluate heuristically the capacity of the model to extract correct information about the value of the parameters that generated the simulated data. We use RJMCMC to estimate the model with a few simulated data examples to illustrate how well the transdimensional estimator performs.

We show the potential importance of including covariates in the model. Covariates can induce overdispersion relative to unconditional binomial variance, which can mislead the estimation. To illustrate the problem we simulate twelve different election scenarios that vary in two dimensions: the proportion of incremental and extreme fraud on the one side and, on the other side, the extent in which there are spatially concentrated covariates that are associated with voter behavior, i.e., with turnout and with the probability to vote for a party. For each combination of values of the parameters we generate 100 simulated data sets with 750 data points each. We label the twelve different election scenarios I-A, II-A, and so on up to IV-C. Using these 12×100 data sets, we estimate the posterior distribution of $\boldsymbol{\theta}$ for each one using diffuse priors as described in section 2.2.1.

We estimate two models, labeled $\mathcal{M}_2$ and $\mathcal{M}_3$. The difference between the models is the inclusion of the covariates in the model: model $\mathcal{M}_2$ includes the covariates, model $\mathcal{M}_3$ does not. When the data are generated by a process in which there is no covariate effect, the two models should agree. When covariates are important and are correctly specified we expect that $\mathcal{M}_2$ provides a posterior distribution concentrated around the correct values of the fraud probabilities. When covariates matter we expect model $\mathcal{M}_3$ to produce incorrect inferences about the fraud probabilities and the expected support for the leading party.

After showing the relevance of the covariates for correct estimation, we show that the MCMC procedure using $\mathcal{M}_2$ has a good performance for different ranges of possible electoral contexts. To do this we provide another series of simulations and sample 20

different combinations of the parameters from their prior. For each one of them, we generate 100 data sets as before and estimate the posterior distribution. For the fixed-dimensional estimator we assess frequentist coverage for key parameters.

3.1 Simulated Data and Estimates

Figure 1 shows empirical bivariate heat plots of the 12 simulated data sets. All 100 data sets for each particular parameter value are pooled in each case. The axes represent the proportion of votes for the leader and the turnout proportion. The panels in each row of the figure represent data produced by the exact same parameters, except for the existence and intensity of the covariate effect. The intensity of the covariate effect here means that we increase the concentration of similar covariates to generate the aggregate counts of votes for the leader and turnout. The columns of the figure represent the exact same covariates, but the proportion of fraud probability in each case is different. So for instance the panels in the first row represent elections in which no fraud has happened. The first column of that row is how the data looks like when there is no covariate effect whatsoever. In the second column of that row, we increased the concentration of the covariates and keep fixed the amount of fraud, which is zero for that first row. Then the third column just intensifies the concentration of units with similar covariates.

*** Figure 1 about here ***

3.1.1 Constant-dimension Models

Table 1 shows the "True" values of the parameters that generated the data in each panel of Figure 1. For instance, data sets II-B plotted in Figure 1 was generated by $\theta_{\text{II-B}}$ whose values of the components are in the respective line in Table 1. That table also presents $\overline{\boldsymbol{\theta}}$ averaged accross all 100 data sets generated for each parameter combination of the models $\mathcal{M}_2$, and $\mathcal{M}_3$. We can see that the estimated values of the posterior expectation for each of the parameters for the model $\mathcal{M}_2$ are very close the the value used to generate the data.

The model $\mathcal{M}_3$ performs as well as $\mathcal{M}_2$ when there is no covariate effect (data sets I-A, II-A, III-A, and IV-A). When there is covariate effect the model $\mathcal{M}_3$ sometimes produces misleading estimated values of the posterior expetation of the fraud probabilities π_I and π_E: incremental fraud probabilities π_I in these cases are worse estimated by $\mathcal{M}_3$ than are extreme fraud probabilities π_E.

*** Table 1 about here ***

Figure 2 shows estimated posterior distributions for each of the 100 data sets generated by parameters values IV-B, described in Table 1. Similar patterns occur for the other eleven data sets. For model $\mathcal{M}_2$ HPD intervals are concentrated around the value of the parameters that generated the data. Frequentist coverage probabilities for $\boldsymbol{\pi}$ for model $\mathcal{M}_2$ are shown in Table 2. The estimated mean (third column) and the HPD interval in the table are averages accross the 100 estimations for each data set. The last two columns present coverage probabilities computed using the intervals for the normal approximation (sixth column, when parameter is not in the boundary) and for the HPD (seventh column). Coverage with the normal approximation intervals is usually conservative: more than 95% of the nominal 95% intervals contain the true value of the parameter. In the last column, we have zero coverage when the values of $\boldsymbol{\pi}$ are on the boundary, as expected (see discussion in Bayarri and Berger (2004); Brown, Cai and DasGupta (2001)), but the averages are very close to the true values of the parameters, and the intervals are very small. Otherwise, the 95% HPD intervals also tend to be conservative. Similar patterns occur for the other parameters, whose figures and tables are presented in supplementary material.

*** Figure 2 and Table 2 about here ***

In sum, the first step establishes that the existence of ignored covariate effects can lead to incorrect estimation of fraud probabilities in a mixture model of binomials, although

ignoring covariates does not necessarily have that consequence. The model that includes correctly specified covariate effects in its structure is capable of capturing the underlying values of the parameters. When the model is correctly specified, inferences about the values of parameters that describe the structure of the model—hence of frauds—should be reasonably accurate.

Figure 3 shows the true values, the estimated values, HPD intervals, and coverage probabilities using the HPD intervals for each of the 20 different values of the parameters. The estimated values are computed by averaging the estimation over the 100 data sets generated for each of the 20 values of the parameters. Figure 3 presents only the estimation for $\boldsymbol{\pi}$, and the values in each panel of the figure are ordered by their true value to facilitate vizualization. Similar figures for the other parameters exist, but they are omitted here. Once again the model $\mathcal{M}_2$ has good performance for different possible combinations of the parameter values.

*** Figure 3 about here ***

3.1.2 Transdimensional Models

In order to assess the ability of the RJMCMC procedure to detect the presence of clusters within the data, simulated data with a known number of clusters were created. For this exercise, the global mixing parameters were the key quantities of interest. However, the values of the other parameters were checked and there was no indication that this procedure created issues with the estimation of the other structural parameters in the model. Convergence is assessed as previously outlined.

For the purposes of simulation, samples of 1000 observations were generated from the logistic-binomial model. The values of π were varied, with particular interest paid to situations where one element of π was close to or equal to zero. 4 chains were run with a burn-in of 10,000 draws and 2,000 draws were taken after the burn-in. In each case, convergence diagnostics were acceptable.

Figure 4 shows one case where each cluster had a significant proportion of observations assigned. This figure demonstrates that the RJMCMC implementation of this model performs similarly to the standard model when each cluster has a large, positive proportion of observations. In each case, the 95% HPD includes the true value.

*** Figure 4 about here ***

To examine the case of interest, runs were done where one of the clusters was set to have 0 observations generated from the known cluster. Figure 5 shows one of these situations. Here both the "No Frauds" and "Extreme Frauds" clusters accurately estimate the known mixing probabilities. Similarly the RJMCMC indicator that shows whether the cluster has positive mass equals one in 100% of the convergent draws. On the other hand "Incremental Fraud" has a 95% HPD that is equal to zero. As can be seen in the indicator's histogram, more than 95% of the draws from the converged chains keep this value at zero. This shows that this model can effectively recognize a cluster with zero mass.

*** Figure 5 about here ***

Another question of interest relates to how well this model deals with small, but positive proportions of observations that belong to a cluster. Figure 6 shows one situation where the extreme fraud cluster is generated to contain 1% of the observations. Like the first model with larger proportions in each cluster, each of the clusters are estimated to have positive probability of occurrence in 100% of draws. This demonstrates the power of this model—even with a relatively small proportion of observations in one of the clusters, the model is able to recognize that the cluster is needed to appropriately model the data.

*** Figure 6 about here ***

A final case worth noting can be seen in Figure 7. Here, the "Incremental Fraud" cluster is generated with only .1% of observations. This shows the limitation of this model.

Though it is known that there is a positive probability of belonging to this cluster, the RJMCMC approach estimates that the probability that it is needed is about .5. As the cluster size approaches zero, the ability of the model to detect that it exists increases in N, but there are still limitations. Regardless, these simulations show that this model is a powerful tool for determining whether or not any fraud exists in the data.

*** Figure 7 about here ***

3.1.3 Observation-level Frauds

Using MCMC evaluations of the models we can assess posterior predictive distributions for the magnitudes of frauds both over the whole electorate and for each aggregation unit observation j. When the model specification is correct the estimated numbers of fraudulent votes are correct when considering the electorate as a whole, but when conditioning on each observation j the estimated number of fraudulent votes for each j can exhibit a small bias. Figure 8 illustrates the approximate unbiasedness of the overall estimate of the number of fraudulent votes in a simulated data set. Not only is the mode of the predictive posterior near the truth but so are the separate totals of "manufactured" and "stolen" votes. Figure 10 shows examples of the small observation-specific biases: the true number of fraudulent votes at each simulated polling station (red lines) is usually a few votes from the predictive posterior mean or HPD interval.

*** Figures 8 and 10 about here ***

Table 4 summarizes biases in the predictive distributions of observation-level frauds for the four simulated datasets ($n = 2000$) and model specifications described in Table 3.[7] RMSE in the table is computed using posterior expectation versus true. Biases are evaluated only for observations that are classified (by Z_i) as fraudulent. Figure 10 graphs the distribution of the observation-level biases in frauds magnitudes.

[7]The bias results in Table 3 are provisional: we need to run the MCMC a little longer.

*** Tables 3 and 4 and Figure 10 about here ***

3.1.4 Losses and Frauds Decisions

We illustrate the use of loss functions and posterior predictive distributions to make decisions about the existence of frauds. We investigate the extent to which having access to dimension-varying technology can be important when making decisions about frauds' existence.

To use (16a) and (16b) to assess frauds at the level of the whole electorate we define Θ_C and Θ_F in terms of the election outcome: was the number of model-fraudulent votes bigger than the margin of victory? We define this margin R as the difference between the first-place and second-place finishers: $R = \sum_{i=1}^{n}(W_i - O_{1i})$ where O_{1i} denotes the count of votes in electoral unit i for the candidate with the second highest number of observed votes overall. In this case $\theta = \sum_{i=1}^{n} F_i$, where F_i denotes the number of model-fraudulent votes at i, $\Theta_C = [0, R)$ and $\Theta_F = [R, \sum_{i=1}^{n} V_i]$ where V_i is the observed count of votes cast at i. To represent a variety of possible valuations of the losses we use $K_C,\ K_F \in \{1, 10, 100, 1000, 10000, 100000, 1000000\}$. Notice that $K_C = K_F = 1$ implies decisions corresponding to those one would make when hypothesis testing.

In four simulated data sets of size $n = 2000$ with the fixed dimension logistic-binomial model (the dataset specifications are reported in Table 3) the decisions do not depend on K_C or K_F: if $\pi = (0.7, 0.3, 0)$ or $\pi = (0.998, 0.001, 0.001)$ then the statistical decision procedure implies frauds should be acted on and actions a_F taken, but if $\pi = (0.5, 0.25, 0.25)$ or $\pi = (0.7, 0, 0.3)$ then actions a_C should be taken. Transdimensional (beta-bernoulli) estimates imply the same decisions.

To use (16a) and (16b) to assess frauds for individual electoral units we use the proportion of fraudulent votes $\theta_i = F_i/V_i$, with $\Theta_{Ci} = [0, .01)$ and $\Theta_{Fi} = [.01, 1]$. Using either the fixed dimensional or a transdimensional logistic-binomial model with simulated data produces identical observation-specific decisions. Table 5 shows the numbers of the

decisions that there are frauds using the data sets described in Table 3. The number of simulated electoral units for which frauds should be acted on does not vary with the magnitudes K_C and K_F.

*** Table 5 about here ***

4 Analysis of Real Data

We used the restricted normal model (without RJMCMC) to estimate fraud probabilities in presidential elections in two countries: Brazil 2014 and Kenya 2013. The electoral systems are different. Brazil has plurality winner-take-all rules and Kenya has plurality rules with a regional dispersion requirement. For Brazil we have polling station or precinct observations and for Kenya we have ward observations. The Brazilian data was aggregated to town level.

Figures 11 shows empirical density plots that display features of the data that these mode-sensitive models should respond to. For the Brazil 2014 election bimodality is neither visually apparent in the marginal densities for the proportions of votes counted for the party with the most votes nor for turnout proportions. For the Kenya 2013 election multimodality is visibly apparent in both the contour plot and the marginal densities.

*** Figure 11 about here ***

The Table 6 shows the posterior means and the 95% high posterior density intervals (HPDI) averaged across all four chains used in the estimation. The Table shows that in Kenya around 70.7% of the votes are non-fraudulent, while 28.9% seems to be result of incremental fraud. These fraud probability estimates agree with the estimates for the Kenyan election in Mebane (2016) using the likelihood version of the Klimek et al. (2012) model. For the Brazilian election, there is no sign of election fraud.

*** Table 6 about here ***

As discussed previously, covariates can affect the shape of the distribution and affect the estimation of fraud. In Brazil, the party that won the 2014 presidential election, the Workers Party (PT), is known to have largest support among the poor population and among people that lives in the northest region of the country. Particularly, PT is associated with a large social program, Bolsa Familia, design to reduce poverty and inequality (Zucco 2010). Therefore, we inclue four covariates and reestimate the model in order to check if their inclusion changes the results. We included the proportion of people receiving Bolsa Familia in each town, the income per capita, the Gini coefficient as a measure of inequality, and the proportion of white people. The Table 7 shows the results. The sign of the coefficients are as expected. PT support is higher among poor and black population, but there is no indication that Bolsa Familia has increased PT support. Moreover the inclusion of the covariates didn't change the previous results, that is, the estimates indicate no fraud in that presidential election.

*** Table 7 about here ***

5 Discussion

As a practical and statistical matter, it is good to shift the positive empirical frauds model framework from a likelihood implementation to a fully Bayesian implementation. Among the advantages a Bayesian implementation confers is the ability to state well-motivated measures of the uncertainty in quantities such as the estimated magnitude of frauds. Predictive distributions from the Bayesian models support making rigorous statistical decisions about whether frauds are present—decisions that take into account the stakes involved in the decisions. So far we have not observed practically consequential differences in performance between the fixed-dimensional and transdimensional approaches, even though model parameter estimates are more accurate when using a transdimensional approach.

Currently we have an operational Bayesian formulations using (1) logistic functions and binomial distributions and (2) restricted Normal distributions. The models and estimators do not exhibit label-switching (e.g. Grün and Leisch 2009). Such an identification concern motivates the link between the two kinds of frauds in the restricted Normal model. Perhaps the good performance of the binomial model means that link can be removed. We have transdimensional MCMC formulations working for the logistic-binomial model, and we don't anticipate difficulties extending the transdimensional approaches to other models.

The logistic-binomial model remains subject to overdispersion concerns. It will hardly ever occur in practice that we have observation-level covariates that we think completely describe how the data were generated. Nonfraudulent election data should be produced from the actions of diverse individuals acting at least somewhat separately, and aggregating the results of those actions into election units like polling stations naturally produces overdispersion (or sometimes underdispersion).

Models based on Normal distributions, that have distinct mean and variance parameters, eliminate such overdispersion concerns. Other approaches are also conceivable, and we are working on alternatives: for example, we're close to having a working beta-binomial model. For interpretability and for the sake of potentially sharp inferences, we think it is important to continue to use parametric model forms. But nonparametric Bayesian models may also be soon in view: the model can identify how many components exist and what their features are; the question will become which if any components correspond to frauds.

Having effective models for "frauds" will still leave the fundamental question of whether the effects of frauds can be distinguished from the effects of individual electors acting strategically. We're developing test beds, using data with simulated strategic behavior and frauds, with which to assess this using the models we are developing.

6 Appendix

6.1 Distributions

The general format of the distributions in equations (5) are obtained by the transformations in (1) and (2). Denote $y = (w, a, i, e) = G(\nu, \tau, \iota, \upsilon)$, and $x = (\nu, \tau, \iota, \upsilon) = G^{-1}(w, a, i, e) = (g_\nu^{-1}, g_\tau^{-1}, g_\iota^{-1}, g_\upsilon^{-1})$, so $G : X \to Y$. Let's restrict the domain of the variables $\nu, \tau, \iota, \upsilon$ to be $(0, 1)^4 \subset \mathbb{R}^4$ so that the jacobian matrix of G^{-1}, denoted J^{-1}, is well defined in the whole domain. Then we know

$$\int_Y f(w, a, i, e) d\mu(y) = \int_X f \circ G^{-1}(w, a, i, e) \left|\det(J^{-1})\right| d\mu(x)$$

In our case $f \circ G^{-1}(w, a, i, e)$ denotes the joint density of $(\nu, \tau, \iota, \upsilon)$. We can derive the density of (w, a) separately for each $Z \in \{1, 2, 3\}$ and in two steps, first deriving $f(a_i \mid \cdot)$, and then $f(w_i, a_i \mid \cdot)$. As before we let $(w, a) = (w_i, a_i)$. We omit the subscript i to simpify the notation. For $Z = 1$ we only need to compute g_ν^{-1} and g_τ^{-1}. We have for a:

$$a = 1 - \tau = g_\tau(\tau) \iff \tau = 1 - a = g_\tau^{-1} \implies \left|\det(J^{-1})\right| = 1$$

$$f(a) = f_\tau(1 - a \mid \mu_\tau, \sigma_\tau) = f(g_\tau^{-1} \mid \mu_\tau, \sigma_\tau)$$

For w we have

$$w = \nu(1 - a) = g_\nu(\nu) \iff \nu = \frac{w}{1 - a} = g_\nu^{-1} \iff \left|\det(J^{-1})\right| = \frac{1}{1 - a}$$

$$f(w) = f(g_\nu^{-1} \mid \mu_\nu, \sigma_\nu) \frac{1}{1 - a}$$

So we have

$$f(w, a) = f(g_\nu^{-1} \mid a, \mu_\nu, \sigma_\nu) f(g_\tau^{-1} \mid \mu_\tau, \sigma_t) \frac{1}{1 - a} = \prod_{j \in \{\nu, \tau\}} f(g_j^{-1} \mid \mu_j, \sigma_j) \frac{1}{1 - a}$$

For $Z = 2$ we have, for a:

$$\begin{cases} a_i = (1-\tau_i)(1-\iota_i^M) = g_\tau \\ i_i^M = \iota_i^M = g_I \end{cases} \quad ; \quad \begin{cases} \tau_i = \left(1 - \dfrac{a_i}{1-\iota_i^M}\right) = g_\tau^{-1} \\ \iota_i^M = i_i^M = g_{\iota^M}^{-1} \end{cases} \quad \implies \left|\det(J^{-1})\right| = \frac{1}{(1-\iota_i^M)}$$

Therefore

$$f(a,i) = f(g_\tau^{-1} \mid \mu_\tau, \sigma_\tau) f(g_{\iota^M}^{-1} \mid \mu_{\iota^M}, \sigma_{\iota^M}) \left(\frac{1}{1-\iota_i^M}\right)$$
$$f(a_i) = \int_{[0,1]} f(g_\tau^{-1} \mid \mu_\tau, \sigma_\tau) f(g_{\iota^M}^{-1} \mid \mu_{\iota^M}, \sigma_{\iota^M}) \left(\frac{1}{1-\iota_i^M}\right) m(d\iota^I)$$

For w_i we have:

$$\begin{cases} w_i = \nu_i \left(\dfrac{1-\iota_i^S}{1-\iota_i^M}\right)\left(1-\iota_i^M - a_i\right) + a_i \left(\dfrac{\iota^M - \iota^S}{1-\iota^M}\right) + \iota^S = g_\nu \\ i_i^M = \iota_i^M = g_{\iota^M} \\ i_i^S = \iota_i^S = g_{\iota^S} \end{cases}$$
$$\begin{cases} \nu_i = w\left(\dfrac{1}{1-\iota_i^M - a_i}\right)\left(\dfrac{1-\iota_i^M}{1-\iota_i^S}\right) - \dfrac{\iota_i^S}{1-\iota_i^S} - \dfrac{a_i \iota_i^M}{(1-\iota_i^M - a_i)(1-\iota_i^S)} = g_\nu^{-1} \\ \iota_i^M = i_i^M = g_{\iota^M}^{-1} \\ \iota_i^S = i_i^S = g_{\iota^S}^{-1} \end{cases}$$

Note that g_ν^{-1} is a straightforward result from the inverse transformation. When $Z_i = 2$,

$\tau_i = 1 - \dfrac{a_i}{1-\iota_i^M}$. Then we have

$$\begin{aligned}
\nu_i &= \left(w_i - \iota_i^M + I_i\tau_i - \iota_i^S\right)\frac{1}{\tau_i(1-\iota_i^S)} \\
&= w_i\left(\frac{1}{\tau_i(1-\iota_i^S)}\right) - \frac{1}{\tau_i}\left(\frac{\iota_i^M}{1-\iota_i^S}\right) + \frac{\iota_i^M}{1-\iota_i^S} - \frac{\iota_i^S}{1-\iota_i^S} \\
&= w_i\left(\frac{1}{\tau_i(1-\iota_i^S)}\right) - \left(\frac{\iota_i^M}{1-\iota_i^S}\right)\left(1-\frac{1}{\tau_i}\right) - \frac{\iota_i^S}{1-\iota_i^S} \\
&= w_i\left(\frac{1}{\tau_i(1-\iota_i^S)}\right) - \left(\frac{\iota_i^M}{1-\iota_i^S}\right)\left(1-\frac{1}{\tau_i}\right) - \frac{\iota_i^S}{1-\iota_i^S} \\
&= w_i\left(\frac{1-\iota_i^M}{1-\iota_i^S}\right)\left(\frac{1}{1-\iota_i^M-a_i}\right) - \frac{\iota_i^S}{1-\iota_i^S} - \left(\frac{\iota_i^M a_i}{(1-\iota_i^S)(1-\iota_i^M-a_i)}\right)
\end{aligned}$$

Then, it is easy to see that $|\det(J^{-1})| = \left(\dfrac{1-\iota_i^M}{1-\iota_i^S}\right)\left(\dfrac{1}{1-\iota_i^M-a_i}\right)$. Therefore

$$\begin{aligned}
f(w_i,a_i) &= \int_{[0,1]^2} f(g_\nu^{-1} \mid \mu_\nu,\sigma_\nu) f(g_\tau^{-1} \mid \mu_\tau,\sigma_\tau) f(g_{\iota^M}^{-1} \mid \mu_{\iota^M},\sigma_{\iota^M}) \\
&\quad \times f(g_{\iota^S}^{-1} \mid \mu_{\iota^S},\sigma_{\iota^S})\left(\frac{1}{(1-\iota_i^S)(1-\iota_i^M-a_i)}\right) m(d\iota^M \times d\iota^S) \\
&= \int_{[0,1]^2} \prod_{j\in\{\nu,\tau,\iota^S,\iota^M\}} f(g_j^{-1} \mid \mu_j,\sigma_j)\left(\frac{1}{(1-\iota_i^M-a_i)(1-\iota_i^S)}\right) m(d\iota^M \times d\iota^S)
\end{aligned}$$

References

Bayarri, M. Jésus and James O. Berger. 2004. "The interplay of Bayesian and Frequentist Analysis." *Statistical Science* 19(1):58–80.

Beber, Bernd and Alexandra Scacco. 2012. "What the Numbers Say: A Digit-Based Test for Election Fraud." *Political Analysis* 20(2):211–234.

Berger, James O. 1985. *Statistical Decision Theory and Bayesian Analysis*. 2 ed. New York: Springer-Verlag.

Blackwell, David and James B MacQueen. 1973. "Ferguson distributions via Pólya urn schemes." *The annals of statistics* pp. 353–355.

Borghesi, C. and J. P. Bouchaud. 2010. "Spatial Correlations in Vote Statistics: A Diffusive Field Model for Decision-making." *European Physical Journal B* 75(3):395–404.

Breunig, Christian and Achim Goerres. 2011. "Searching for Electoral Irregularities in an Established Democracy: Applying Benford's Law Tests to Bundestag Elections in Unified Germany." *Electoral Studies* 30(3):534–545.

Brooks, Steve, Andrew Gelman, Galin Jones and Xiao-Li Meng, eds. 2011. *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: CRC Press.

Brown, Lawrence D., T. Tony Cai and Anirban DasGupta. 2001. "Interval Estimation for a Binomial Proportion." *Statistical Science* 16(1):101–117.

Cantu, Francisco and Sebastián M. Saiegh. 2011. "Fraudulent Democracy? An Analysis of Argentina's Infamous Decade Using Supervised Machine Learning." *Political Analysis* 19(4):409–433.

Carlin, John B. and Thomas A. Louis. 2009. *Bayesian Data Analysis*. 3 ed. Boca Raton, FL: Chapman & Hall/CRC.

Cowles, Mary Katherine and Bradley P. Carlin. 1996. "Markov Chain Monte Carlo Convergence Diagnostics: a Comparative Review." *Journal of the American Statistical Association* 91(434):883–904.

de Valpine, P., D. Turek, C.J. Paciorek, C. Anderson-Bergman, D. Temple Lang and R.

Bodik. 2017. "Programming with models: writing statistical algorithms for general model structures with NIMBLE." *Journal of Computational and Graphical Statistics* 26:403–413.

Deckert, Joseph, Mikhail Myagkov and Peter C. Ordeshook. 2011. "Benford's Law and the Detection of Election Fraud." *Political Analysis* 19(3):245–268.

Dunson, David B and Ju-Hyun Park. 2008. "Kernel stick-breaking processes." *Biometrika* 95(2):307–323.

Ferguson, Thomas S. 1973. "A Bayesian analysis of some nonparametric problems." *The annals of statistics* pp. 209–230.

Flegal, James M. 2008. "Monte Carlo Standard Errors for Markov Chain Monte Carlo." PhD thesis, University of Minnesota.

Flegal, James M. and J. Hughes. 2012. "mcmcse: Monte Carlo Standard Errors for MCMC." Riverside, CA and Minneapolis, MN. R package version.

Flegal, James M., Murali Haran and Galin L Jones. 2008. "Markov Chain Monte Carlo: Can We Trust the Third Significant Figure?" *Statistical Science* 23(2):250–260.

Gelman, Andrew and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7(4):457–472.

Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin. 2014. *Bayesian Data Analysis*. 2 ed. Boca Raton, FL: Chapman & Hall/CRC.

Gelman, Andrew and Kenneth Shirley. 2011. Inference From Simulations and Monitoring Convergence. In *Handbook of Markov Chain Monte Carlo*, ed. Steve Brooks, Andrew Gelman, Galin Jones and Xiao-Li Meng. Boca Raton, FL: CRC Press pp. 163–174.

Gelman, Andrew and Xiao-Li Meng. 1998. "Simulating normalizing constants: From importance sampling to bridge sampling to path sampling." *Statistical science* pp. 163–185.

Gershman, Samuel J and David M Blei. 2012. "A tutorial on Bayesian nonparametric models." *Journal of Mathematical Psychology* 56(1):1–12.

Gong, Lei and James M. Flegal. 2015. "A Practical Sequential Stopping Rule for High-dimensional Markov Chain Monte Carlo." *Journal of Computational and Graphical Statis-*

tics forthcoming(.):–.

Green, Peter J. 1995. "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination." *Biometrika* 82(4):711–732.

Grün, Bettina and Friedrich Leisch. 2009. "Dealing with Label Switching in Mixture Models Under Genuine Multimodality." *Journal of Multivariate Analysis* 100:851–861.

Hicken, Allen and Walter R. Mebane, Jr. 2015. "A Guide to Election Forensics." Working paper for IIE/USAID subaward #DFG-10-APS-UM, "Development of an Election Forensics Toolkit: Using Subnational Data to Detect Anomalies".

Hjort, Nils Lid. 1990. "Nonparametric Bayes estimators based on beta processes in models for life history data." *The Annals of Statistics* pp. 1259–1294.

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

Jones, Galin L., Murali Haran, Brian S. Caffo and Ronald Neath. 2006. "Fixed-width Output Analysis for Markov Chain Monte Carlo." *Journal of the American Statistical Association* 101(476):1537–1547.

Klimek, Peter, Yuri Yegorov, Rudolf Hanel and Stefan Thurner. 2012. "Statistical Detection of Systematic Election Irregularities." *Proceedings of the National Academy of Sciences* 109(41):16469–16473.

Levin, Inès, Gabe A. Cohn, Peter C. Ordeshook and R. Michael Alvarez. 2009. "Detecting Voter Fraud in an Electronic Voting Context: An Analysis of the Unlimited Reelection Vote in Venezuela." 2009 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections, Montreal, August 10–11.

Mebane, Jr., Walter R. 2010. "Fraud in the 2009 Presidential Election in Iran?" *Chance* 23:6–15.

Mebane, Jr., Walter R. 2015. "Election Forensics Toolkit DRG Center Working Paper." Working paper for IIE/USAID subaward #DFG-10-APS-UM, "Development of an Election Forensics Toolkit: Using Subnational Data to Detect Anomalies".

Mebane, Jr., Walter R. 2016. "Election Forensics: Frauds Tests and Observation-level Frauds Probabilities." Paper presented at the 2016 Annual Meeting of the Midwest Political Science Association, Chicago, April 7–10, 2016.

Meyn, Sean P and Richard L. Tweedie. 2012. *Markov Chains and Stochastic Stability*. New York: Springer Science & Business Media.

Montgomery, Jacob M., Santiago Olivella, Joshua D. Potter and Brian F. Crisp. 2015. "An Informed Forensics Approach to Detecting Vote Irregularities." *Political Analysis* forthcoming.

Myagkov, Mikhail, Peter C. Ordeshook and Dimitry Shaikin. 2009. *The Forensics of Election Fraud: With Applications to Russia and Ukraine*. New York: Cambridge University Press.

Myagkov, Misha, Peter C. Ordeshook and Dimitry Shaikin. 2008. Estimating the Trail of Votes in Russia's Elections and the Likelihood of Fraud. In *The Art and Science of Studying Election Fraud: Detection, Prevention, and Consequences*, ed. R. Michael Alvarez, Thad E. Hall and Susan D. Hyde. Washington, DC: Brookings Institution.

Paisley, John and Lawrence Carin. 2009. Nonparametric factor analysis with beta process priors. In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM pp. 777–784.

Paisley, John and Michael I Jordan. 2016. "A constructive definition of the beta process." *arXiv preprint arXiv:1604.00685* .

Paisley, John W, Aimee K Zaas, Christopher W Woods, Geoffrey S Ginsburg and Lawrence Carin. 2010. A stick-breaking construction of the beta process. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. pp. 847–854.

Paisley, John W, David M Blei and Michael I Jordan. 2012. Stick-breaking beta processes and the Poisson process. In *International Conference on Artificial Intelligence and Statistics*. pp. 850–858.

Pericchi, Luis Raúl and David Torres. 2011. "Quick Anomaly Detection by the Newcomb-Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico

and Venezuela." *Statistical Science* 26(4):502–516.

Pitman, Jim et al. 2002. "Combinatorial stochastic processes.".

Plummer, Martyn. 2003. "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.".

Plummer, Martyn. 2010. *JAGS: A Program for Analysis of Bayesian Graphical Models*. Version 2.2.0.

Plummer, Martyn, Alexey Stukalov and Matt Denwood. 2016. "rjags: Bayesian Graphical Models using MCMC." URL `https://cran.r-project.org/web/packages/rjags/index.html`, linked to JAGS 4.2.0.

Rasmussen, Carl Edward. 2000. The infinite Gaussian mixture model. In *Advances in neural information processing systems*. pp. 554–560.

Robert, Christian and George Casella. 2013. *Monte Carlo Statistical Methods*. New York: Springer Science & Business Media.

Rozenas, Arturas. 2017. "Detecting Election Fraud from Irregularities in Vote-Share Distributions." *Political Analysis* 25(1):41–56.

Shikano, Susumu and Verena Mack. 2009. "When Does the Second-Digit Benford's Law-Test Signal an Election Fraud? Facts or Misleading Test Results." *Jahrbücher fur Nationalökonomie und Statistik* 231(5-6):719–732.

Teh, Yee Whye. 2011. Dirichlet process. In *Encyclopedia of machine learning*. Springer pp. 280–287.

Vats, Dootika, James M. Flegal and Galin L. Jones. 2015. "Multivariate Output Analysis for Markov chain Monte Carlo." arXiv:1512.07713.

Wand, Jonathan, Kenneth Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr., Michael Herron and Henry E. Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95:793–810.

Zucco, Cesar. 2010. "Cash-transfers and voting behavior: An assessment of the political impacts of the Bolsa Família program." *Draft. Available at http://www. princeton. edu/\ ˜*

zucco/papers/cctvotingbehavior.[Accessed March 21, 2011] .

Table 1: Estimation for each data set and model

Data	Model	π_0	π_I	π_E	β_0	β_1	β_2	γ_0	γ_1	γ_2	ζ^a	ζ^o	χ^a	χ^o
I-A	True	1.0000	0.0000	0.0000	0.3000	0.0000	0.0000	-0.3000	0.0000	0.0000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.9992	0.0004	0.0004	0.2979			-0.2998			0.3658	0.3803	0.8902	0.8792
	$\mathcal{M}_3$	0.9992	0.0004	0.0004	0.2979			-0.2999			0.3898	0.2820	0.8092	0.8123
I-B	True	1.0000	0.0000	0.0000	0.3000	0.0500	0.3000	-0.3000	1.0000	-0.4000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.9992	0.0004	0.0004	0.2959	0.0474	0.3011	-0.2993	1.0001	-0.3990	0.3886	0.1338	0.9217	0.8637
	$\mathcal{M}_3$	0.6378	0.3528	0.0094	-0.0450			0.1601			0.5390	0.0000	0.7595	0.7004
I-C	True	1.0000	0.0000	0.0000	0.3000	0.0500	0.3000	-0.3000	1.0000	-0.4000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.9990	0.0006	0.0004	0.3035	0.0509	0.3004	-0.3000	0.9986	-0.4004	0.1770	0.1358	0.8732	0.8657
	$\mathcal{M}_3$	0.6185	0.3658	0.0157	-0.0921			0.2076			0.5485	0.0000	0.7169	0.7002
II-A	True	0.8500	0.1500	0.0000	0.3000	0.0000	0.0000	-0.3000	0.0000	0.0000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.8671	0.1325	0.0004	0.2959			-0.3001			0.1975	0.2130	0.8680	0.9007
	$\mathcal{M}_3$	0.8672	0.1324	0.0004	0.2961			-0.3000			0.1978	0.2119	0.8677	0.8069
II-B	True	0.8500	0.1500	0.0000	0.3000	0.0500	0.3000	-0.3000	1.0000	-0.4000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.8675	0.1321	0.0004	0.3012	0.0513	0.3025	-0.3002	1.0000	-0.4008	0.1957	0.2049	0.8651	0.8955
	$\mathcal{M}_3$	0.6213	0.3633	0.0154	0.0360			0.1285			0.5370	0.0000	0.7410	0.7002
II-C	True	0.8500	0.1500	0.0000	0.3000	0.0500	0.3000	-0.3000	1.0000	-0.4000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.8651	0.1345	0.0004	0.3034	0.0469	0.2965	-0.3030	1.0015	-0.3984	0.2011	0.2067	0.9097	0.9456
	$\mathcal{M}_3$	0.5945	0.3858	0.0197	-0.0207			0.1856			0.5498	0.0000	0.7155	0.7002
III-A	True	0.8000	0.0000	0.2000	0.3000	0.0000	0.0000	-0.3000	0.0000	0.0000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.7811	0.0004	0.2185	0.2947			-0.2993			0.4595	0.4531	0.9003	0.9000
	$\mathcal{M}_3$	0.7811	0.0004	0.2185	0.2946			-0.2993			0.3748	0.3175	0.9003	0.9002
III-B	True	0.8000	0.0000	0.2000	0.3000	0.0500	0.3000	-0.3000	1.0000	-0.4000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.7620	0.0195	0.2185	0.2986	0.0484	0.3002	-0.3003	1.0012	-0.3985	0.1225	0.0313	0.9010	0.9013
	$\mathcal{M}_3$	0.5081	0.2720	0.2199	-0.0202			0.1485			0.5510	0.0000	0.9207	0.8951
III-C	True	0.8000	0.0000	0.2000	0.3000	0.0500	0.3000	-0.3000	1.0000	-0.4000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.7920	0.0006	0.2074	0.2989	0.0506	0.2964	-0.3001	1.0018	-0.3983	0.1936	0.0543	0.9011	0.9042
	$\mathcal{M}_3$	0.5078	0.2849	0.2074	-0.0674			0.1931			0.5669	0.0000	0.9237	0.9005
IV-A	True	0.7000	0.2000	0.1000	0.3000	0.0000	0.0000	-0.3000	0.0000	0.0000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.6761	0.2164	0.1075	0.3019			-0.3005			0.2051	0.1965	0.9019	0.8992
	$\mathcal{M}_3$	0.6761	0.2164	0.1075	0.3019			-0.3006			0.2049	0.1966	0.9019	0.8991
IV-B	True	0.7000	0.2000	0.1000	0.3000	0.0500	0.3000	-0.3000	1.0000	-0.4000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.6776	0.2149	0.1075	0.3012	0.0503	0.3026	-0.2994	0.9977	-0.3992	0.2017	0.1949	0.9018	0.9022
	$\mathcal{M}_3$	0.5461	0.3452	0.1088	0.1190			0.0974			0.5509	0.0000	0.9187	0.8871
IV-C	True	0.7000	0.2000	0.1000	0.3000	0.0500	0.3000	-0.3000	1.0000	-0.4000	0.2000	0.2000	0.9000	0.9000
	$\mathcal{M}_2$	0.7019	0.1910	0.1071	0.2917	0.0382	0.3072	-0.3018	0.9978	-0.4017	0.2031	0.1977	0.9005	0.8997
	$\mathcal{M}_3$	0.5455	0.3468	0.1077	0.0570			0.1422			0.5712	0.0000	0.9206	0.8936

Table 2: Estimation and Coverage Probability for $\boldsymbol{\phi}$ for estimation using model $\mathcal{M}_2$

dataset	Parameter	True	Estimated.Mean	Aver.HPD	Coverage.SD	Coverage.HPD
I-A-a	π_0	1.0000	0.9960	(0.9904, 0.9999)	1	0
	π_I	0.0000	0.0020	(0, 0.0061)	1	0
	π_E	0.0000	0.0020	(0, 0.006)	1	0
I-B-a	π_0	1.0000	0.9656	(0.9117, 0.9999)	0.98	0
	π_I	0.0000	0.0324	(0, 0.0851)	0.98	0
	π_E	0.0000	0.0020	(0, 0.006)	1	0
I-C-a	π_0	1.0000	0.9559	(0.9329, 0.9999)	0.96	0
	π_I	0.0000	0.0421	(0, 0.0639)	0.96	0
	π_E	0.0000	0.0020	(0, 0.006)	1	0
II-A-a	π_0	0.8500	0.8502	(0.8186, 0.8809)	0.92	0.92
	π_I	0.1500	0.1478	(0.1173, 0.1792)	0.92	0.92
	π_E	0.0000	0.0020	(0, 0.0059)	1	1
II-B-a	π_0	0.8500	0.8510	(0.8193, 0.8818)	0.94	0.94
	π_I	0.1500	0.1471	(0.1163, 0.1785)	0.94	0.94
	π_E	0.0000	0.0020	(0, 0.006)	1	1
II-C-a	π_0	0.8500	0.8463	(0.8141, 0.8775)	0.96	0.94
	π_I	0.1500	0.1517	(0.1205, 0.1835)	0.96	0.96
	π_E	0.0000	0.0020	(0, 0.006)	1	1
III-A-a	π_0	0.8000	0.8003	(0.7651, 0.8349)	0.96	0.96
	π_I	0.0000	0.0020	(0, 0.0065)	1	0
	π_E	0.2000	0.1976	(0.1634, 0.2326)	0.96	0.96
III-B-a	π_0	0.8000	0.7836	(0.7084, 0.8345)	1	0.98
	π_I	0.0000	0.0188	(0, 0.0664)	1	0
	π_E	0.2000	0.1976	(0.1633, 0.2325)	0.96	0.96
III-C-a	π_0	0.8000	0.7774	(0.7106, 0.8301)	0.98	0.98
	π_I	0.0000	0.0198	(0, 0.0581)	1	0
	π_E	0.2000	0.2028	(0.1681, 0.2381)	0.96	0.96
IV-A-a	π_0	0.7000	0.6975	(0.6574, 0.7375)	0.96	0.96
	π_I	0.2000	0.2041	(0.1694, 0.2397)	0.98	0.98
	π_E	0.1000	0.0984	(0.073, 0.1246)	0.98	0.96
IV-B-a	π_0	0.7000	0.6989	(0.6583, 0.7387)	0.94	0.94
	π_I	0.2000	0.2028	(0.1679, 0.2385)	0.98	0.98
	π_E	0.1000	0.0984	(0.073, 0.1246)	0.98	0.96
IV-C-a	π_0	0.7000	0.6978	(0.6573, 0.7379)	0.98	0.98
	π_I	0.2000	0.2004	(0.1657, 0.236)	0.98	1
	π_E	0.1000	0.1018	(0.0762, 0.1285)	0.94	0.94

Table 3: Observation-level Frauds Simulation Datasets Parameters

Parameter	True	Posterior Mean	95% HPD (lower)	95% HPD (upper)
Data Set 1				
β_1	0.2394	0.2604	0.2140	0.2991
β_2	−0.0009	0.0019	−0.0048	0.0095
γ_1	−0.0694	−0.0728	−0.0914	−0.0523
γ_2	0.1213	0.1231	0.1191	0.1274
μ_υ^M	0.9956	0.9928	0.9924	0.9932
μ_υ^S	0.8403	0.8900	0.8873	0.8926
μ_ι^M	0.3872	0.3871	0.3833	0.3909
μ_ι^S	0.0986	0.0978	0.0854	0.1134
π_1	0.5000	0.5184	0.4966	0.5403
π_2	0.2500	0.2427	0.2246	0.2618
π_3	0.2500	0.2389	0.2203	0.2573
Data Set 2				
β_1	−0.2219	−0.2504	−0.2841	−0.2135
β_2	0.0794	0.0792	0.0726	0.0850
γ_1	−0.0481	−0.0467	−0.0668	−0.0278
γ_2	−0.1622	−0.1655	−0.1692	−0.1620
μ_υ^M	0.8949	0.8324	0.8036	0.8520
μ_υ^S	0.9910	0.8570	0.8191	0.9017
μ_ι^M	0.4247	0.4257	0.4227	0.4287
μ_ι^S	0.2358	0.2355	0.2289	0.2422
π_1	0.7000	0.6999	0.6792	0.7194
π_2	0.3000	0.2996	0.2804	0.3208
π_3	0.0000	0.0005	0.0000	0.0015
Data Set 3				
β_1	0.0927	0.1000	0.0652	0.1299
β_2	−0.0497	−0.0466	−0.0527	−0.0402
γ_1	0.2401	0.2494	0.2325	0.2677
γ_2	−0.2218	−0.2244	−0.2286	−0.2207
μ_υ^M	0.8758	0.8756	0.8739	0.8775
μ_υ^S	0.9158	0.9160	0.9124	0.9195
μ_ι^M	0.1210	0.3614	0.3327	0.3868
μ_ι^S	0.1691	0.3392	0.2829	0.3796
π_1	0.7000	0.6949	0.6743	0.7143
π_2	0.0000	0.0005	0.0000	0.0015
π_3	0.3000	0.3046	0.2851	0.3251
Data Set 4				
β_1	−0.0453	−0.0389	−0.0660	−0.0086
β_2	0.1388	0.1379	0.1324	0.1437
γ_1	−0.0883	−0.0734	−0.0882	−0.0597
γ_2	0.2352	0.2326	0.2292	0.2360
μ_υ^M	0.9682	0.9370	0.8968	0.9602
μ_υ^S	0.8716	0.8890	0.8643	0.9190
μ_ι^M	0.0068	0.3403	0.2944	0.3788
μ_ι^S	0.1124	0.3550	0.3312	0.3803
π_1	0.9980	0.9980	0.9960	0.9996
π_2	0.0010	0.0005	0.0000	0.0015
π_3	0.0010	0.0015	0.0002	0.0032

Table 4: Observation-level Fraud Magnitudes Bias Statisitcs

	Bias			
Fraud	Average	Minimum	Maximum	RMSE
Data Set 1				
Total Fraud Bias	2.0231	−44.0981	32.0098	10.1743
Manufactured Fraud Bias	2.0231	−44.0981	32.0098	10.1743
Stolen Fraud Bias	2.0231	−44.0981	32.0098	10.1743
Data Set 2				
Total Fraud Bias	0.3051	−32.2843	32.5867	7.7904
Manufactured Fraud Bias	0.3051	−32.2843	32.5867	7.7904
Stolen Fraud Bias	0.3051	−32.2843	32.5867	7.7904
Data Set 3				
Total Fraud Bias	0.6127	−35.1933	33.6829	10.4103
Manufactured Fraud Bias	0.6127	−35.1933	33.6829	10.4103
Stolen Fraud Bias	0.6127	−35.1933	33.6829	10.4103
Data Set 4				
Total Fraud Bias	−17.5418	−51.0654	14.7852	30.4214
Manufactured Fraud Bias	−17.5418	−51.0654	14.7852	30.4214
Stolen Fraud Bias	−17.5418	−51.0654	14.7852	30.4214

Table 5: Observation-level Frauds Decision Counts, by Loss Specifications

Data Set 1		K_F						
		1	10	100	1000	10000	1e+05	1e+06
	1	963	963	963	963	963	963	963
	10	963	963	963	963	963	963	963
	100	963	963	963	963	963	963	963
K_C	1000	963	963	963	963	963	963	963
	10000	963	963	963	963	963	963	963
	1e+05	963	963	963	963	963	963	963
	1e+06	963	963	963	963	963	963	963

Data Set 2		K_F						
		1	10	100	1000	10000	1e+05	1e+06
	1	599	599	599	599	599	599	599
	10	599	599	599	599	599	599	599
	100	599	599	599	599	599	599	599
K_C	1000	599	599	599	599	599	599	599
	10000	599	599	599	599	599	599	599
	1e+05	599	599	599	599	599	599	599
	1e+06	599	599	599	599	599	599	599

Data Set 3		K_F						
		1	10	100	1000	10000	1e+05	1e+06
	1	609	609	609	609	609	609	609
	10	609	609	609	609	609	609	609
	100	609	609	609	609	609	609	609
K_C	1000	609	609	609	609	609	609	609
	10000	609	609	609	609	609	609	609
	1e+05	609	609	609	609	609	609	609
	1e+06	609	609	609	609	609	609	609

Data Set 4		K_F						
		1	10	100	1000	10000	1e+05	1e+06
	1	2	2	2	2	2	2	2
	10	2	2	2	2	2	2	2
	100	2	2	2	2	2	2	2
K_C	1000	2	2	2	2	2	2	2
	10000	2	2	2	2	2	2	2
	1e+05	2	2	2	2	2	2	2
	1e+06	2	2	2	2	2	2	2

Note: fixed-dimension binomial-logistic models or transdimensional (beta-bernoulli) binomial-logistic models; sample size $n = 2000$. Data Set 1, $\pi = (0.5, 0.25, 0.25)$; Data Set 2, $\pi = (0.7, 0.3, 0)$; Data Set 3, $\pi = (0.7, 0, 0.3)$; Data Set 4, $\pi = (0.998, 0.001, 0.001)$.

	Kenya 2013		
Parameter	Mean	95% HPD.lower	95% HPD.upper
No Fraud (π_0)	0.7072	0.6633	0.7798
Incremental Fraud (π_I)	0.2893	0.2072	0.3354
Extreme Fraud (π_E)	0.0034	0.0000	0.0200
β_ν	−19.3493	−25.7663	−7.1961
β_τ	0.1680	0.0563	0.2364
α	0.1756	0.0546	0.7032
μ_χ^M	0.7908	0.7021	0.9695
μ_ι^M	0.1394	0.0016	0.4458
σ_ι^M	0.2211	0.1560	0.2883
σ_ν	2.8692	1.9257	3.3707
σ_τ	0.3746	0.3366	0.4202

	Brazil (2014)		
Parameter	Mean	95% HPD.lower	95% HPD.upper
No Fraud (π_0)	0.9996	0.9988	1.0000
Incremental Fraud (π_I)	0.0002	0.0000	0.0009
Extreme Fraud (π_E)	0.0002	0.0000	0.0007
β_ν	0.6067	0.5987	0.6144
β_τ	0.7352	0.7325	0.7381
α	1.1279	0.0325	3.9107
μ_χ^M	0.9267	0.8100	0.9964
μ_ι^M	0.3390	0.0177	0.6841
σ_ι^M	1.1389	0.1889	3.5530
σ_ν	0.2431	0.2054	0.2548
σ_τ	0.0961	0.0925	0.1106

Table 6: Estimates of the restricted normal model for Brazil and Kenya (column "Mean" represents the posterior average)

	Brazil (2014)		
Parameter	Mean	95% HPD.lower	95% HPD.upper
No Fraud (π_I)	0.9996	0.9987	0.9999
Incremental Fraud (π_I)	0.0003	0.0000	0.0011
Extreme Fraud (π_E)	0.0002	0.0000	0.0007
Coefficients for proportion of Votes for the winner			
Intercept ($\beta_{0\nu}$)	0.7124	0.6482	0.7740
Proportion of White ($\beta_{1\nu}$)	−0.0562	−0.1352	−0.0074
Inequality ($\beta_{2\nu}$)	0.0777	0.0002	0.2155
Bolsa Familia ($\beta_{3\nu}$)	0.0000	0.0000	0.0000
Income per capita ($\beta_{4\nu}$)	−0.0008	−0.0010	−0.0007
Coefficients for proportion of abstention			
Intercept ($\beta_{0\tau}$)	0.6878	0.3236	0.7647
Proportion of White ($\beta_{1\tau}$)	0.1184	0.0672	0.3101
Inequality ($\beta_{2\tau}$)	0.0205	−0.0841	0.6383
Bolsa Familia ($\beta_{3\tau}$)	−0.0000	−0.0000	0.0000
Income per capita ($\beta_{4\tau}$)	−0.0000	−0.0001	0.0001
α	1.1903	0.0339	4.0329
μ_χ^M	0.8288	0.7074	0.9771
μ_ι^M	0.2979	0.0141	0.6714
σ_ι^M	0.9731	0.1537	3.5561
σ_ν	0.1895	0.1446	0.2223
σ_τ	0.1311	0.1081	0.1923

Table 7: Estimates of the restricted normal model that includes covariate effect for Brazil (column "Mean" represents the posterior average)

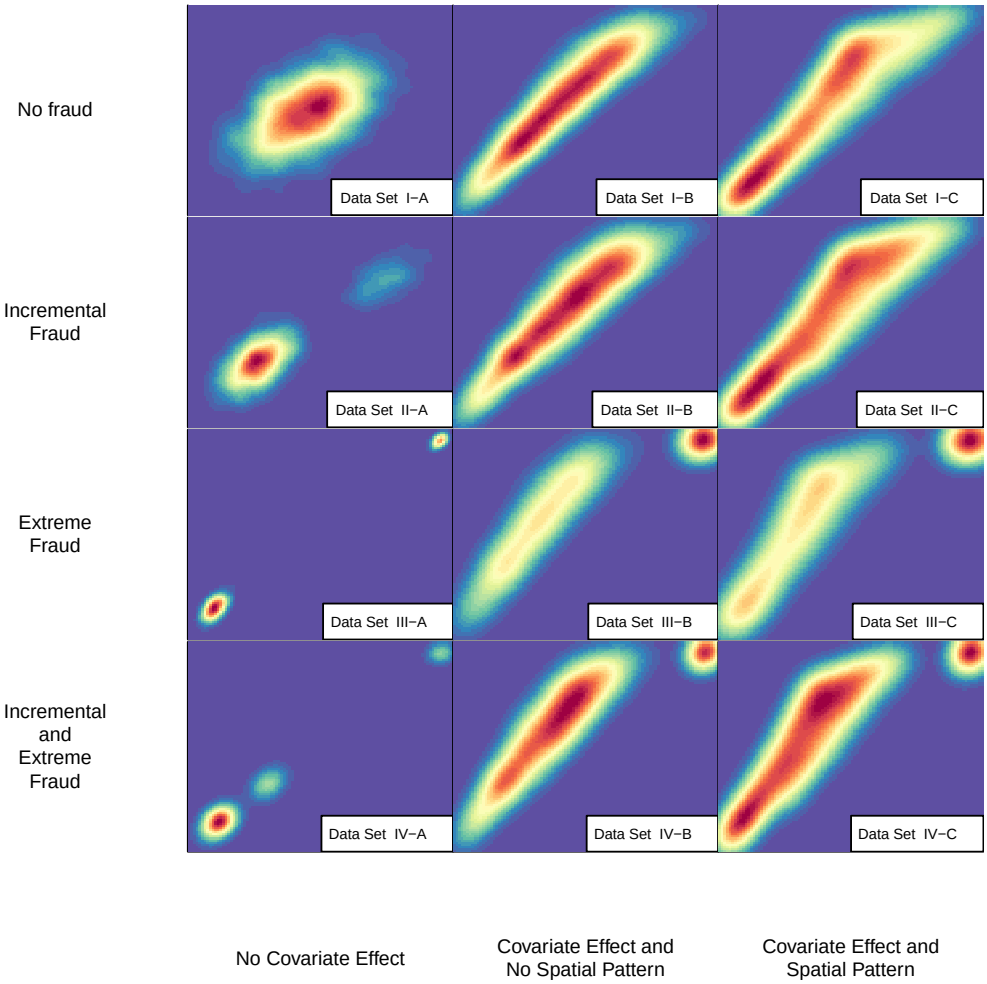


Figure 1: Contour plots of the bivariate distribution of vote proportions for the leader (w_i) and turnout $t_i = 1 - a_i$ from simluated data sets using parameters of Table 1.

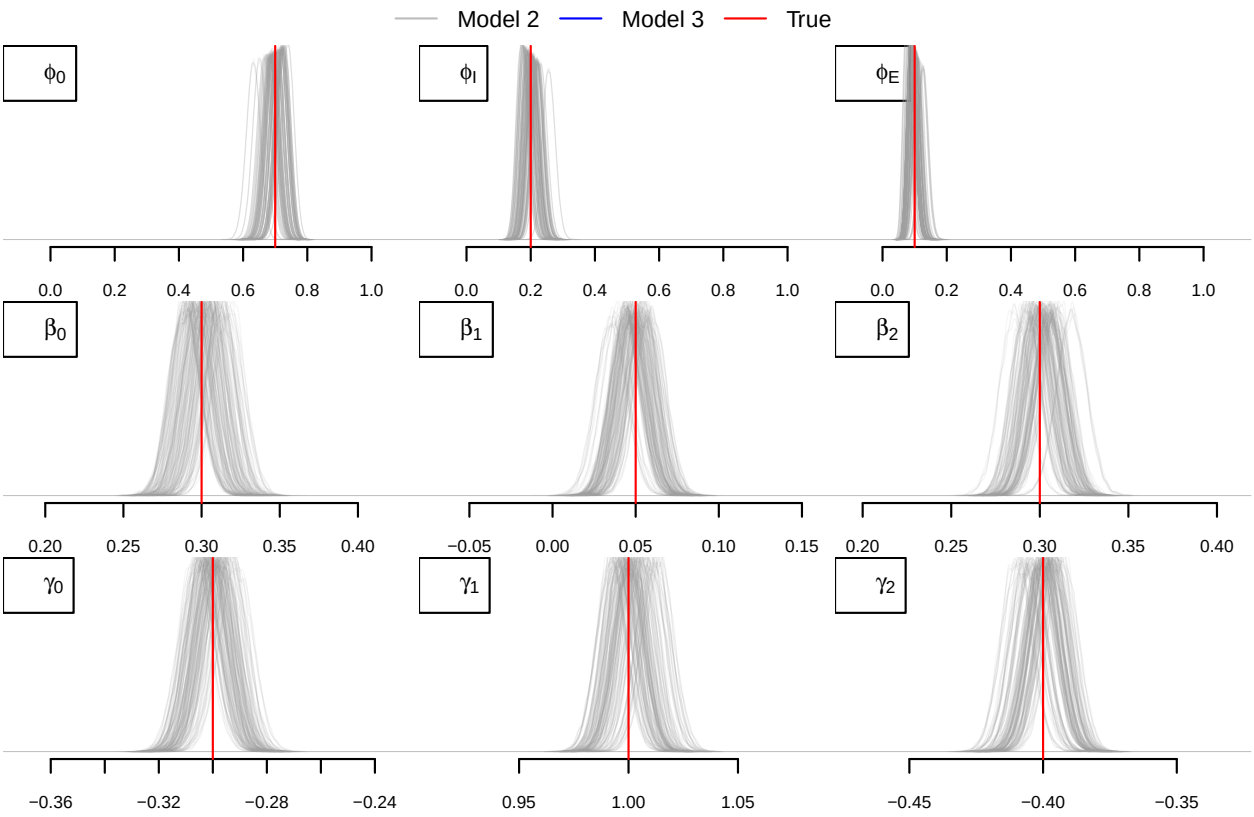


Figure 2: Posterior distribution of 100 simulated data sets using θ_{IV-B}, that is, an election with both incremental and extreme fraud, and moderate spatial concentration of the covariates.

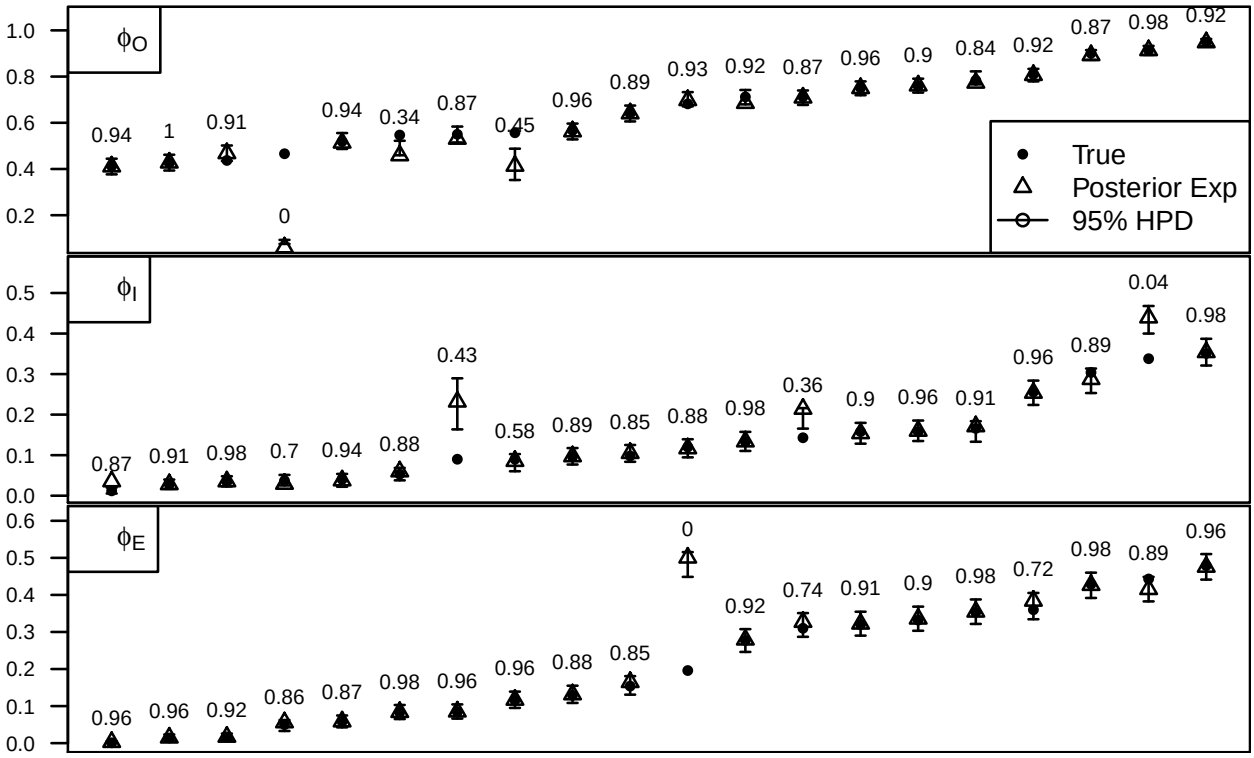


Figure 3: Estimation for each one of the 20 different parameter vector averaged across 100 data sets. Values shown in increasing order of true value for each parameter. Numbers represent the HPD coverage probability.

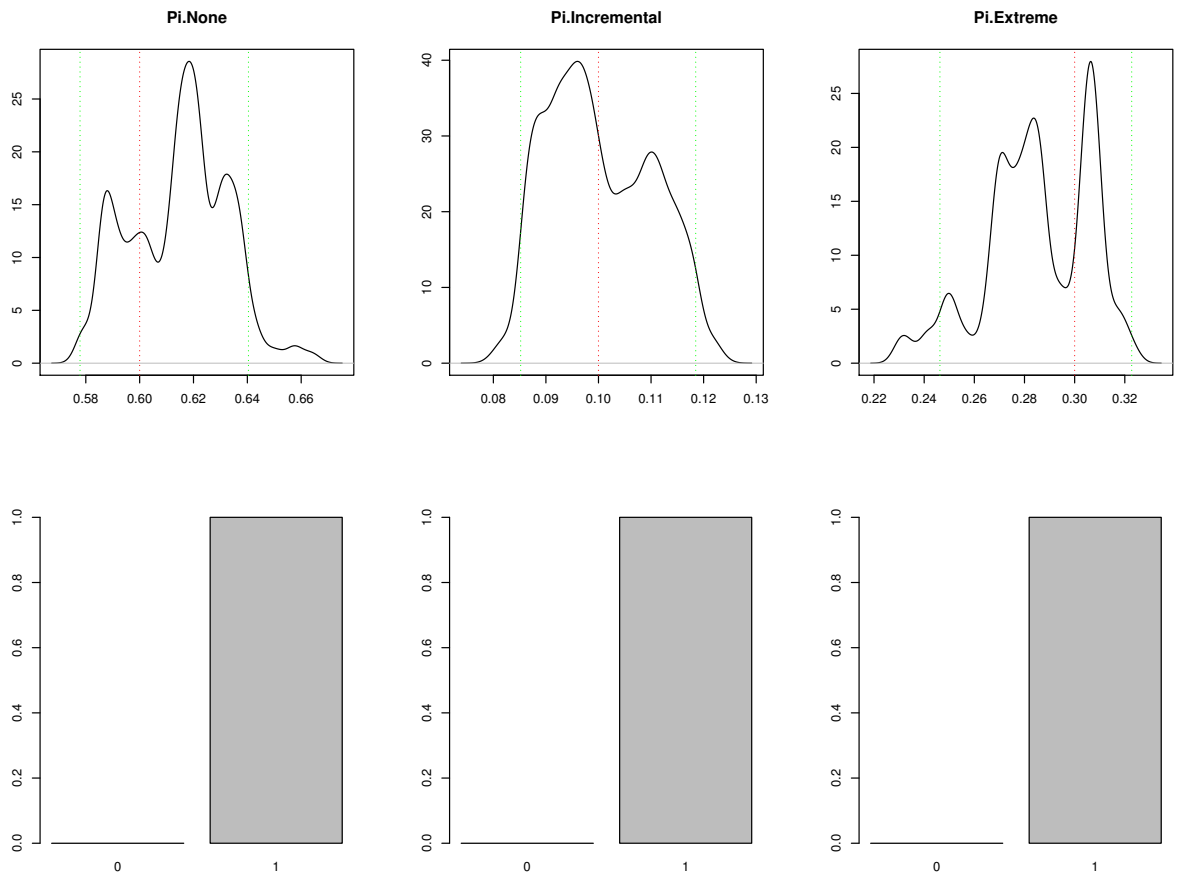


Figure 4: Density Plots of RJMCMC Global Mixing Parameters for Full Model with Known Positive Probability in Each Cluster

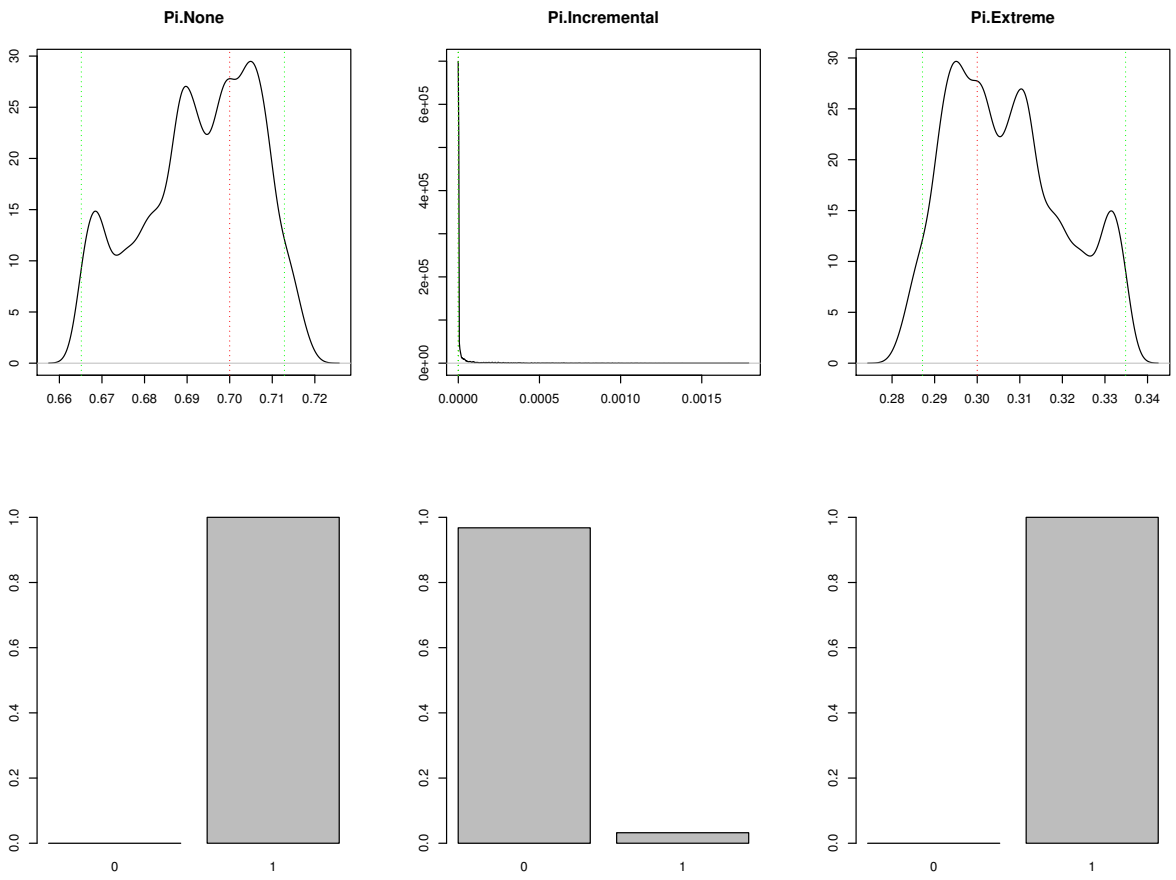


Figure 5: Density Plots of RJMCMC Global Mixing Parameters for Full Model with Known Positive Probability in Two Clusters and Zero Probability in the Third

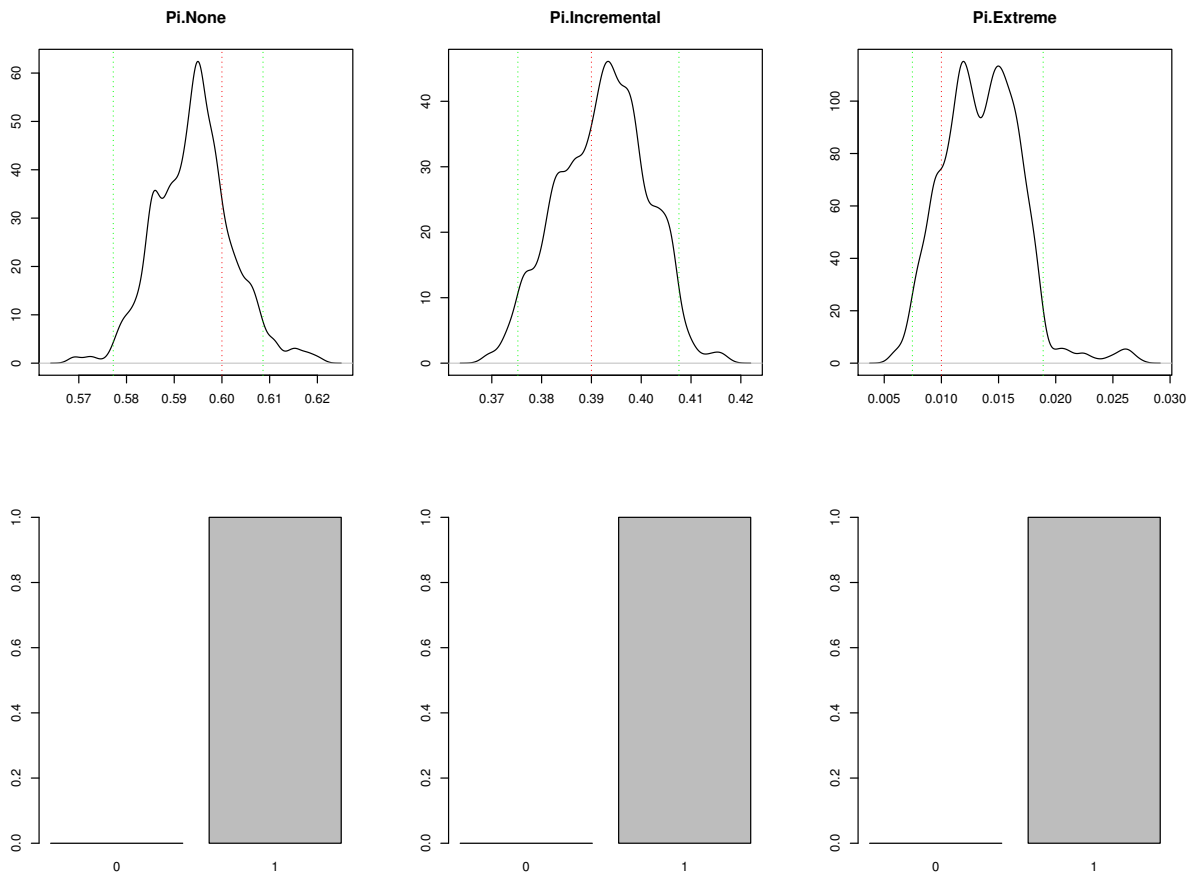


Figure 6: Density Plots of RJMCMC Global Mixing Parameters for Full Model with Known Positive Probability in Two Clusters and Small (but Positive) probability in the Third

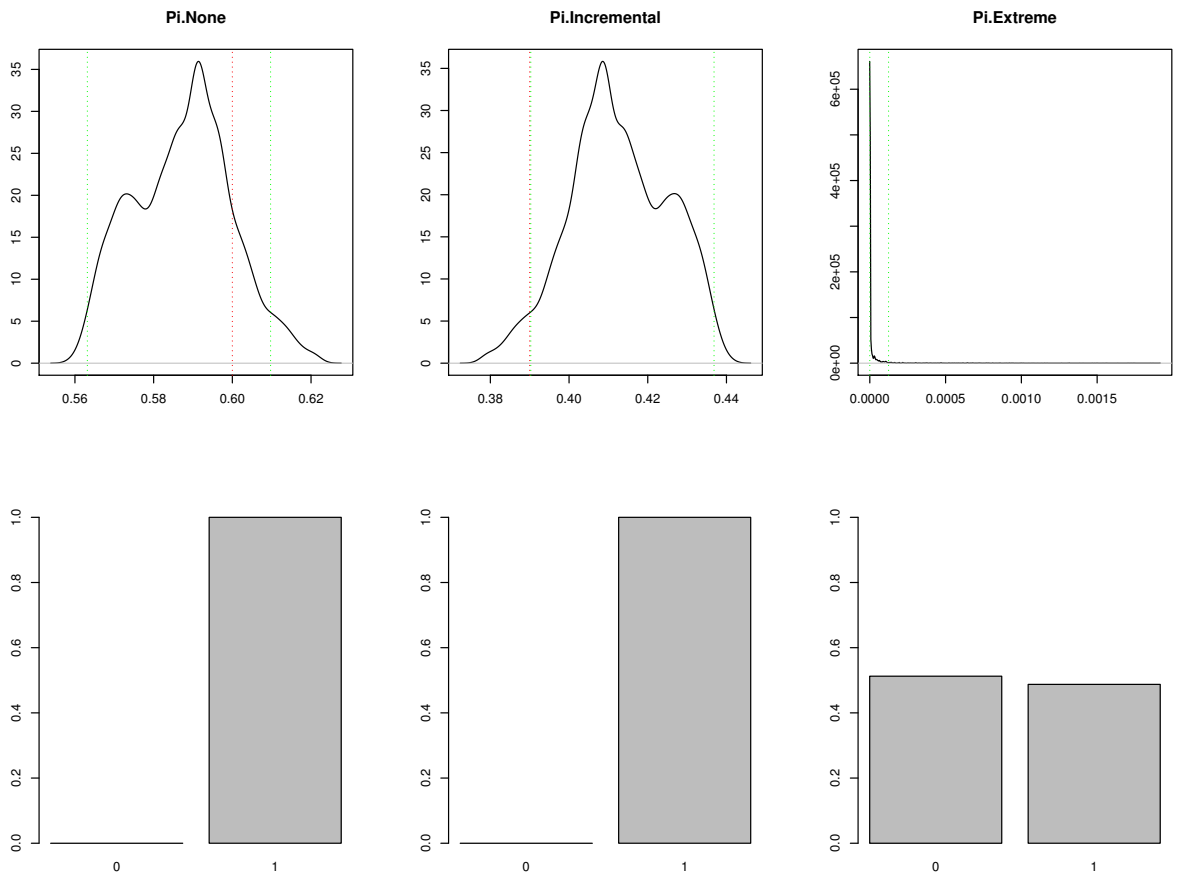


Figure 7: Density Plots of RJMCMC Global Mixing Parameters for Full Model with Known Positive Probability in Two Clusters and Extremely Small (but Positive) probability in the Third

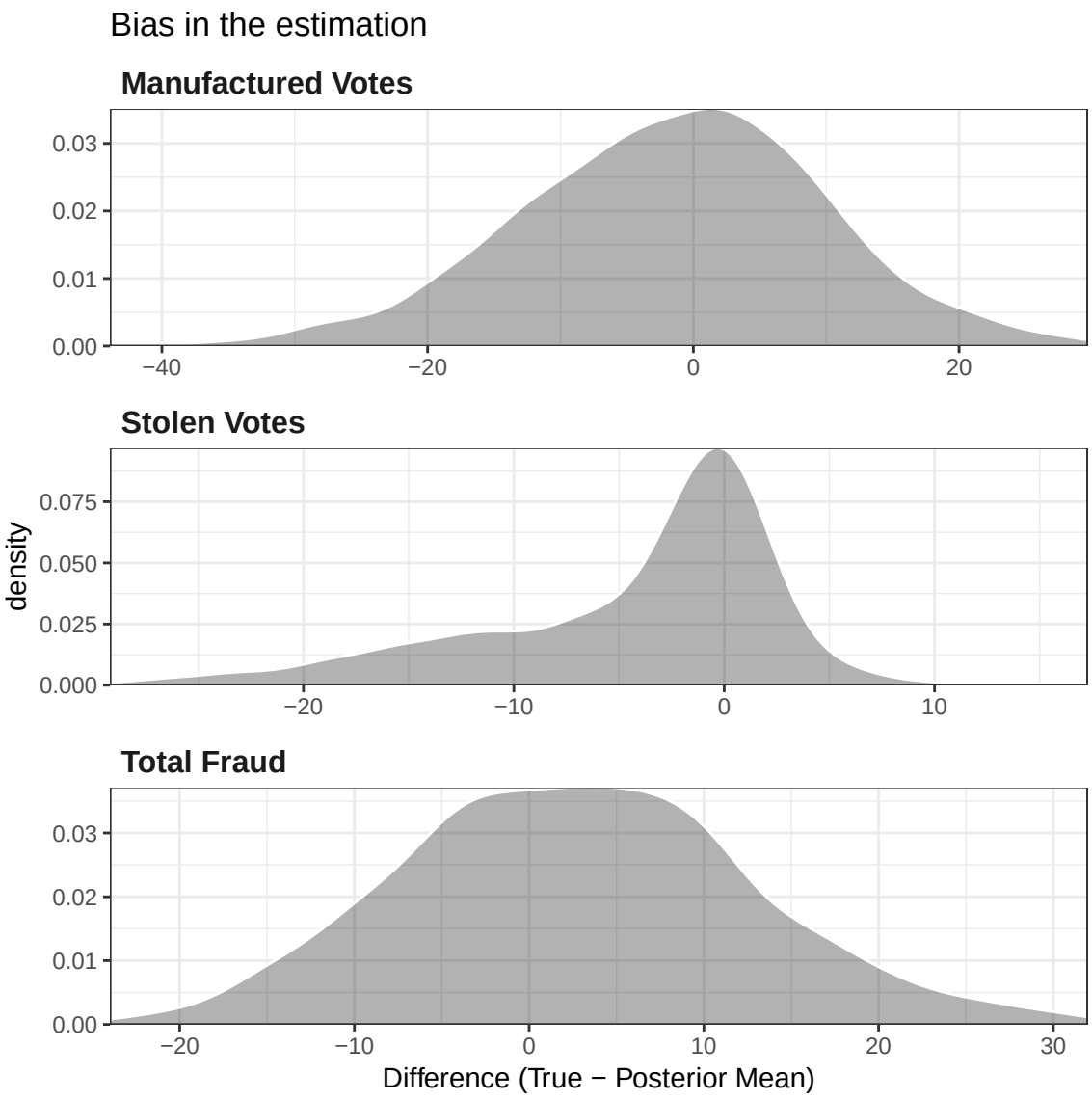


Figure 8: Overall Fraud Magnitude Predictive Distribution Example

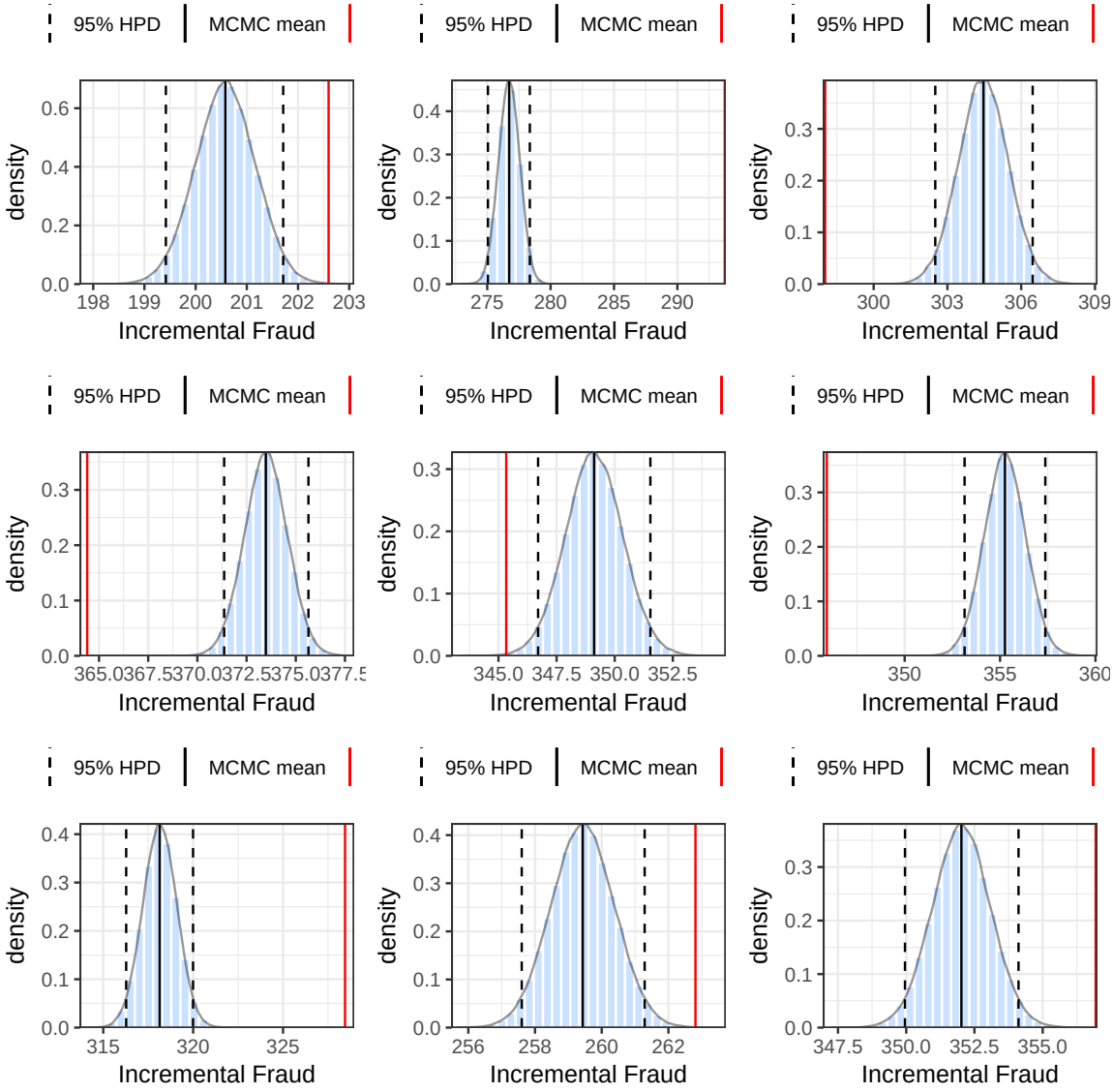


Figure 9: Observation-level Fraud Magnitude Predictive Distribution Examples

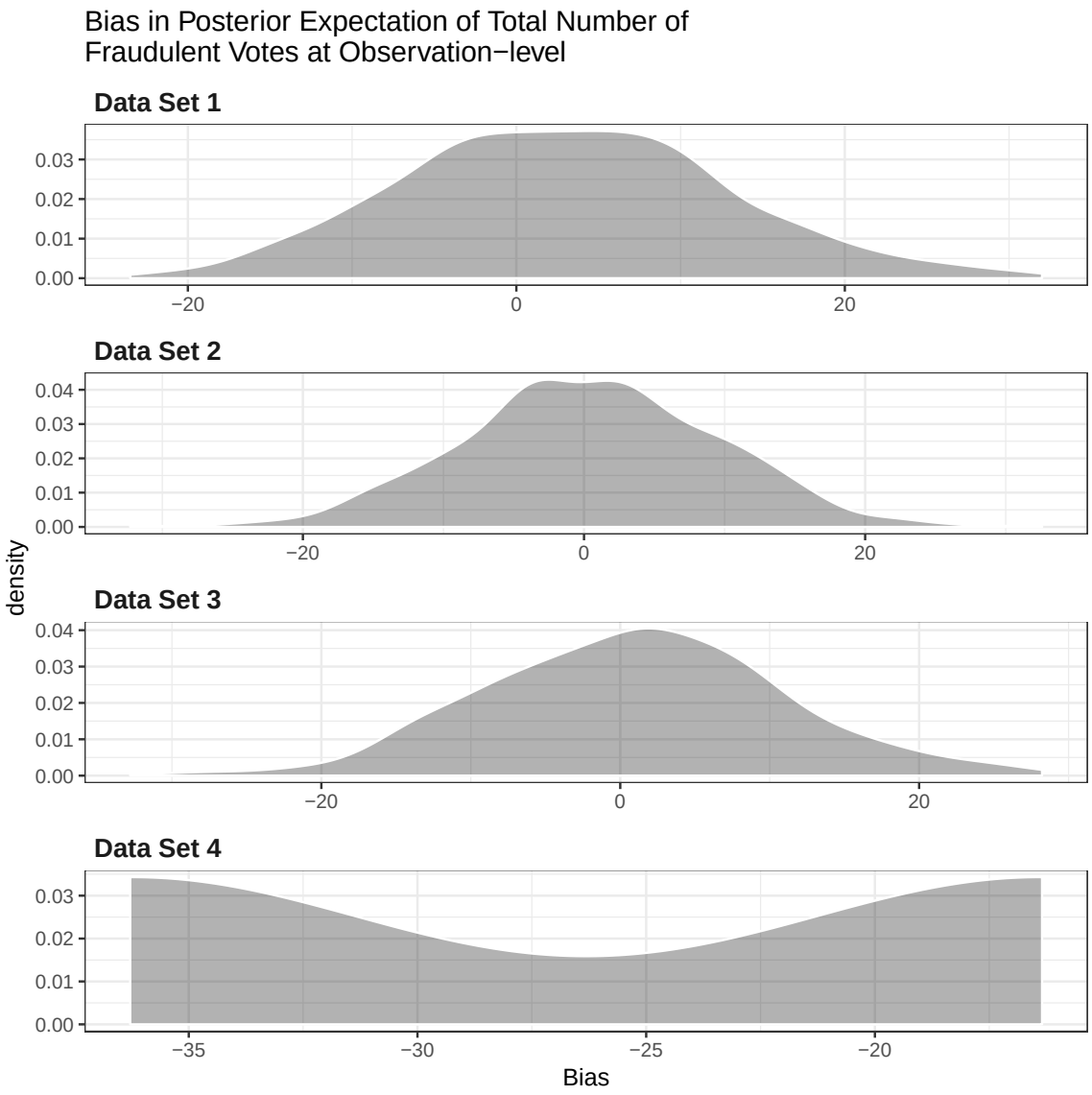
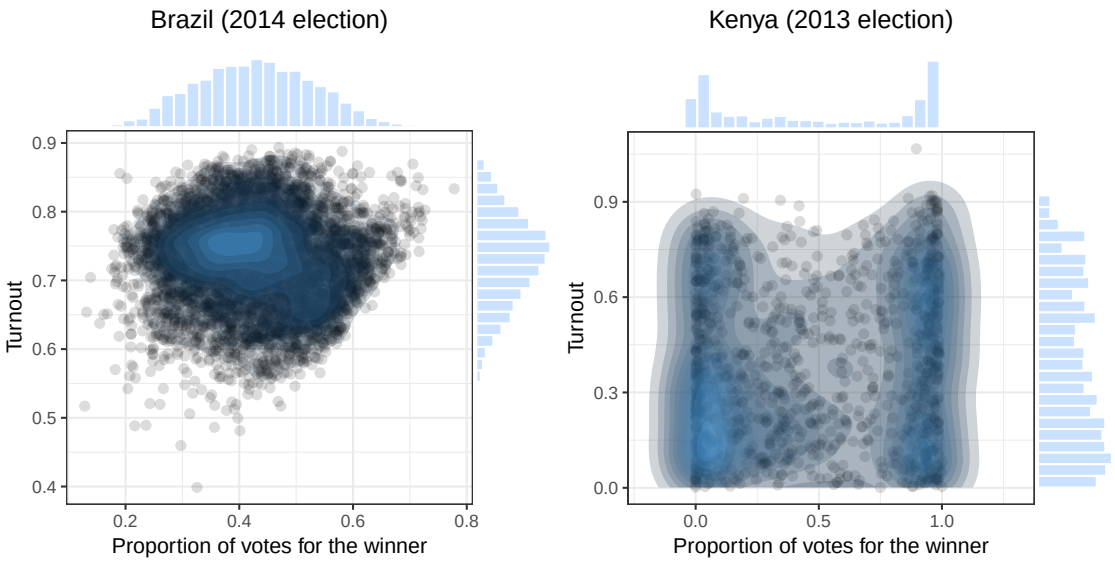


Figure 10: Observation-level Fraud Magnitude Biases



Note: $n = 5567$ town observations for Brazil; $n = 1,231$ ward observations for Kenya.

Figure 11: Empirical densities for Brazil (left) and Kenya (right)