Developments in Positive Empirical Models of Election Frauds*

Diogo Ferrari† Walter R. Mebane, Jr.‡

July 2, 2017

*Prepared for presentation at the 2017 Summer Meeting of the Political Methodology Society, Madison, WI, July 13–15, 2017. 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 and at the 2017 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 6–9, 2017. Work supported by NSF award SES 1523355. Thanks to Kevin McAlister for suggestions.

†Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: dferrari@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 develop a general Bayesian framework using finite mixture models of product distributions to identify the probability and distribution of frauds in elections. The framework is based on transformations of random variables that capture incremental and extreme frauds inspired by the work of Klimek et al (2012). A formulation that closely matches their specifications is a special case. The framework can be used with a larger class of probability distributions (i.e. parametric and distributional assumptions) than the original model that includes a mixture of restricted normal distributions. The general formulation explicitly models effects of covariates associated with voters' behavior and with the occurrence of frauds. We present logistic-binomial and restricted Normal variants of the general model. We present an extension intended to handle absentee/mail precincts, where we do not observe how many electors acted in each type of election unit. We use simulated data to show that the logistic-binomial model can be estimated well using MCMC methods and that the estimates have good frequentist coverage properties when the model is correctly specified. Estimates are not as good when relevant covariates are ignored.

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). 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 also describe modifications required to handle absentee (or mail) ballots that have a particular organization that frequently occurs.

We adopt a Bayesian approach to formulating the model, which we estimate using Markov Chain Monte Carlo (MCMC) methods (Plummer 2003; Plummer, Stukalov and Denwood 2016). The model allows 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. We report some of those results, including investigation of the frequentist coverage properties of the Bayesian estimates produced using MCMC methods.

2 Model Motivation

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 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:

[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.

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

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

3. no fraud (occurs with probability $f_0 = 1 - f_{\rm i} - f_{\rm e}$): 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 $f_{\rm i}$): $x_i \sim |\mathcal{N}(0, \theta)|$, $0 < x_i < 1$, 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 $f_{\rm e}$): 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).

3 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:

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

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

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

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

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

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. We also introduce an extension to accomodate absentee ballots that have a particular structure.

4 The Logistic-Binomial Finite Mixture Model

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

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

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:

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

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

$$\pi \sim \text{Dirichlet}(\sigma_1, \sigma_2, \sigma_3), \quad \sigma_1, \sigma_2, \sigma_3 \in R_+$$

The assumptions on the dependence structure of the random variables in the model are, $\forall i, j = 1, ..., N$ and $i \neq j$, $(A1)(\nu_i \perp\!\!\!\perp \nu_j) \mid X$, $(A2)(A_i \perp\!\!\!\perp A_j) \mid X$, $(A3)(\nu_i \perp\!\!\!\perp A_j) \mid X$, $(A4)S_i^l \perp\!\!\!\perp E_i^l \perp\!\!\!\perp Z_i$ 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.

5 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.

[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, $\iota_i^l \mid X_i \sim f(i_i^l \mid \mu_\iota^l, \sigma_\iota^l)$, $i_i^l \in [0, 1]$.

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 I and E in the sampling distribution as expressed by the equations in (5), as was necessary in Mebane (2016). By conditioning the sample distribution on I and E, 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 Bayesian Restricted Normal Finite Mixture (BRNFM) model:

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

where

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

and

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

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, E_i, and I_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

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

6 Mail/Absentee Precincts

Sometimes there is a mix of "in-person" and "absentee" (or "mail") polling stations, where each absentee polling station comprises a set of the in-person polling stations, and each voter is free to choose whether to vote in-person or absentee (or not vote).[4] The dilemma is that observed data tells us how many electors are eligible to act at each type of polling station but not how many do. We observe the numbers of votes cast at each type of polling station but not the numbers of electors who considered voting at each but did not: we do not observe the counts of abstentions A_i.

Each in-person polling station is associated with at most one absentee station; some in-person stations may not be associated with an absentee station. The number of eligible voters in each in-person unit is observed as N_i, $i = 1, \ldots, n_0$, where n_0 is the number of in-person polling stations. The set of in-person polling stations associated with absentee station j is $\mathfrak{m}_j$, $j = n_0 + 1, \ldots, n_0 + n_1$. Each absentee station has M_j eligible voters, with $M_j = \sum_{i \in \mathfrak{m}_j} N_i$ being the sum of the counts of eligible voters in $\mathfrak{m}_j$. The total number of polling stations is $n = n_0 + n_1$.

For each $i = 1, \ldots, n$ 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 parties (the "opposition"), denoted O_i. The number of valid votes is observed as $V_i = W_i + O_i$, $i = 1, \ldots, n$.

The question is what observables to associate with true turnout τ_i, which might be framed as how to characterize the number of nonvotes A_i. It does not work to say that abstentions for in-person stations are $A_i = N_i - V_i$, $i = 1, \ldots, n_0$, because some who did

[4]Examples include election data from Germany, from Austria, from California and from other U.S. states in various years.

not vote in-person may have voted absentee. Nor does it work to say that $A_j = M_j - V_j$, $j = n_0 + 1, \ldots, n_0 + n_1$, because many who did not vote absentee may have voted in person.

Let φ_i, $i = 1, \ldots, n_0$, be the unknown proportion of electors at i who act in person instead of absentee. Then abstention numbers are

$$A_i = \begin{cases} N_i\varphi_i - V_i, \quad i = 1, \ldots, n_0 \\ \sum_{k \in \mathfrak{m}_i} N_k(1 - \varphi_k) - V_i, \quad i = n_0 + 1, \ldots, n_0 + n_1\,. \end{cases} \tag{13}$$

"Observed" turnout is

$$t_i = \begin{cases} \dfrac{V_i}{N_i\varphi_i}\,, \quad i = 1, \ldots, n_0 \\ \dfrac{V_i}{\sum_{k \in \mathfrak{m}_i} N_k(1 - \varphi_k)}\,, \quad i = n_0 + 1, \ldots, n_0 + n_1\,. \end{cases} \tag{14}$$

When we don't know φ_i, these turnout proportions are not in fact observed.

The unknown φ_i, $i = 1, \ldots, n_0$, are to be estimated.[5] These terms may be functions of covariates. We can use A_i and t_i terms based on (13) and (14) in modified models based on (1) with $a_i = 1 - t_i$: for $i = 1, \ldots, n_0$ and for $j = n_0 + 1, \ldots, n_0 + n_1$,

$$N_i\varphi_i - V_i = \begin{cases} N_i\varphi_i(1 - \tau_i) & \text{, if } Z_i = 1 \\ N_i\varphi_i(1 - \tau_i)(1 - \iota_i^M) & \text{, if } Z_i = 2 \\ N_i\varphi_i(1 - \tau_i)(1 - \upsilon_i^M) & \text{, if } Z_i = 3 \end{cases} \tag{15a}$$

$$\sum_{i \in \mathfrak{m}_j} N_i(1 - \varphi_i) - V_j = \begin{cases} \left(\sum_{i \in \mathfrak{m}_j} N_i(1 - \varphi_i)\right)(1 - \tau_j) & \text{, if } Z_i = 1 \\ \left(\sum_{i \in \mathfrak{m}_j} N_i(1 - \varphi_i)\right)(1 - \tau_j)(1 - \iota_j^M) & \text{, if } Z_i = 2 \\ \left(\sum_{i \in \mathfrak{m}_j} N_i(1 - \varphi_i)\right)(1 - \tau_j)(1 - \upsilon_j^M) & \text{, if } Z_i = 3 \end{cases} \tag{15b}$$

For estimation purposes (15) can be rearranged to isolate V_i and V_j on the lefthand sides.

[5] If i is not in an absentee polling station set, then $\varphi_i = 1$.

Equation (2) is unchanged. We must add structure to support estimating the unknown φ_i. Whether φ_i, τ_i, $i \in \mathfrak{m}_j$, and τ_j are distinctly identifiable is a question. Because $N_i\varphi_i \geq V_i$ for $i = 1, \ldots, n_0$, lower bounds are given by $\varphi_i \geq V_i/N_i$. Upper bounds are given by solutions to $\sum_{i\in\mathfrak{m}_j} N_i(1-\varphi_i) \geq V_j$, for $j = n_0+1, \ldots, n_0+n_1$, which generally does not include $\forall i, \varphi_i = 1$.[6]

If there might be fraud in the absentee ballot process, things are more complicated.

Sketch of a More General Absentee Frauds Model: The idea of fraud with absentee/mail ballot counts having the structure we stipulate here may involve new kinds of frauds, which we call "cross-level" frauds. There are four possibilities: (1) abstention at the in-person ballot can be fabricated as vote for the leader in the absentee ballot; (2) vote for the opposition in i can be counted as vote for the leader in m_j; (3) abstention or (4) vote for opposition in m_j counted as in-person vote for the leader in i. We model the magnitudes of these possibilities in terms of proportions

λ_{ji}^a : absentee abstention at j counted as in-person vote for leader in $i \in \mathfrak{m}_j$

λ_{ji}^o : absentee vote for opposition at j counted as in-person vote for leader in $i \in \mathfrak{m}_j$

κ_{ij}^a : in-person abstention in $i \in \mathfrak{m}_j$ counted as absentee vote for leader at j

κ_{ij}^o : in-person vote for opposition in $i \in \mathfrak{m}_j$ counted as absentee vote for leader at j

where $\sum_{i\in\mathfrak{m}_j} \lambda_{ji}^a \leq 1$ and $\sum_{i\in\mathfrak{m}_j} \lambda_{ji}^o \leq 1$.

An example of specifications we might use for abstention as defined in (13) follows.

[6] If $\mathfrak{m}_j$ contains only one in-person station i, then $1 - \dfrac{V_j}{N_i} \geq \varphi_i$ is an upper bound. If $\mathfrak{m}_j$ contains more than one in-person station, then $\sum_{i\in\mathfrak{m}_j} N_i(1-\varphi_i) \geq V_j$ does not simplify except to express one proportion given all the other proportions:

$$1 - \frac{V_j}{N_{i_1}} + \frac{\sum_{i\neq i_1 \& i\in\mathfrak{m}_j}(1-\varphi_i)N_i}{N_{i_1}} \geq \varphi_{i_1}$$

is an upper bound.

Given $i \in \mathfrak{m}_j$, for $i = 1, \ldots, n_0$,

$$A_i = \begin{cases} N_i\varphi_i(1-\tau_i) & \text{, if } Z_i = 1 \\ N_i\varphi_i(1-\tau_i)(1-\kappa_{ij}^a)(1-\iota_i^M) & \text{, if } Z_i = 2 \\ N_i\varphi_i(1-\tau_i)(1-\kappa_{ij}^a)(1-\upsilon_i^M) & \text{, if } Z_i = 3 \end{cases}$$

and for $j = n_0 + 1, \ldots, n_0 + n_1$,

$$A_j = \begin{cases} \left(\sum_{i\in\mathfrak{m}_j} N_i(1-\varphi_i)\right)(1-\tau_j) & \text{, if } Z_j = 1 \\ \left(\sum_{i\in\mathfrak{m}_j} N_i(1-\varphi_i)\right)(1-\tau_j)\left(1-\sum_{i\in\mathfrak{m}_j}\lambda_{ji}^a\right)(1-\iota_j^M) & \text{, if } Z_j = 2 \\ \left(\sum_{i\in\mathfrak{m}_j} N_i(1-\varphi_i)\right)(1-\tau_j)\left(1-\sum_{i\in\mathfrak{m}_j}\lambda_{ji}^a\right)(1-\upsilon_i^M) & \text{, if } Z_j = 3 \end{cases}$$

Such a specification might be too simple from a conceptual standpoint. For instance, frauds within a polling station may not always coincide with frauds between in-person and absentee stations. And more distinct magnitudes of frauds might be plausible, while the preceding specifications have the same magnitudes of cross-level frauds occurring with both type $Z = 2$ and type $Z = 3$ frauds.

Specifications can also be formulated for the votes for the leader. We do not describe those here.

The idea of cross-level frauds is rich, but we have serious concerns about whether the idea leads to models that are feasible to estimate. There may be too many latent random variables, dwarfing the identification problems we already face with the φ_i parameters. Before starting serious work on the more complicated models, we'll need to learn how well the simpler specification described by (2) and (15) can capture motivating suspected fraudulent activity.

7 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 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 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 4.

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. We assess frequentist coverage for key parameters.

7.1 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 $[\Sigma]_{kk} = \sigma_{\theta_k} = \sqrt{\mathbb{V}\mathrm{ar}[\theta_{k1}] + 2\sum_{i=2}^{\infty} \mathbb{C}\mathrm{ov}[\theta_{k1}, \theta_{ki}]}$ 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.

7.2 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 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 ***

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 ϕ_I and ϕ_E: incremental fraud probabilities ϕ_I in these cases are worse estimated by $\mathcal{M}_3$ than are extreme fraud probabilities ϕ_E.

***** Table 1 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 $\boldsymbol{\phi}$ 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 $\boldsymbol{\phi}$ 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 HERE *****

***** Table 2 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 $\boldsymbol{\phi}$, 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 HERE *****

8 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 should confer is the ability to state well-motivated measures of the uncertainty in quantities such as the estimated magnitude of frauds. Currently we have an operational Bayesian formulation using logistic functions and binomial distributions. The model and estimator does not exhibit label-switching (cf. 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.

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, will eliminate such overdispersion concerns. Other approaches are also conceivable. For interpretability and for the sake of potentially sharp inferences, we think it is important to continue to use parametric model forms. We'll soon begin work along these lines.

We're still tweaking and testing the restricted Normal model and just beginning work on extended models like the model for absentee/mail precincts.

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.

9 Appendix

9.1 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:

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

Therefore

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

For w_i we have:

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

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

Note that g_ν^{-1} is a straightforward result from the inverse transformation. When $Z_i = 2$, $\tau_i = 1 - \frac{a_i}{1-\iota_i^M}$. 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 $|\det(J^{-1})| = \left(\frac{1-\iota_i^M}{1-\iota_i^S}\right)\left(\frac{1}{1-\iota_i^M - a_i}\right)$. Therefore

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

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.

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.

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

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.

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

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".

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.

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.

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

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.

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

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.

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.

Table 1: Estimation for each data set and model

Data	Model	ϕ_O	ϕ_I	ϕ_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	ϕ_0	1.0000	0.9960	(0.9904, 0.9999)	1	0
	ϕ_I	0.0000	0.0020	(0, 0.0061)	1	0
	ϕ_E	0.0000	0.0020	(0, 0.006)	1	0
I-B-a	ϕ_0	1.0000	0.9656	(0.9117, 0.9999)	0.98	0
	ϕ_I	0.0000	0.0324	(0, 0.0851)	0.98	0
	ϕ_E	0.0000	0.0020	(0, 0.006)	1	0
I-C-a	ϕ_0	1.0000	0.9559	(0.9329, 0.9999)	0.96	0
	ϕ_I	0.0000	0.0421	(0, 0.0639)	0.96	0
	ϕ_E	0.0000	0.0020	(0, 0.006)	1	0
II-A-a	ϕ_0	0.8500	0.8502	(0.8186, 0.8809)	0.92	0.92
	ϕ_I	0.1500	0.1478	(0.1173, 0.1792)	0.92	0.92
	ϕ_E	0.0000	0.0020	(0, 0.0059)	1	1
II-B-a	ϕ_0	0.8500	0.8510	(0.8193, 0.8818)	0.94	0.94
	ϕ_I	0.1500	0.1471	(0.1163, 0.1785)	0.94	0.94
	ϕ_E	0.0000	0.0020	(0, 0.006)	1	1
II-C-a	ϕ_0	0.8500	0.8463	(0.8141, 0.8775)	0.96	0.94
	ϕ_I	0.1500	0.1517	(0.1205, 0.1835)	0.96	0.96
	ϕ_E	0.0000	0.0020	(0, 0.006)	1	1
III-A-a	ϕ_0	0.8000	0.8003	(0.7651, 0.8349)	0.96	0.96
	ϕ_I	0.0000	0.0020	(0, 0.0065)	1	0
	ϕ_E	0.2000	0.1976	(0.1634, 0.2326)	0.96	0.96
III-B-a	ϕ_0	0.8000	0.7836	(0.7084, 0.8345)	1	0.98
	ϕ_I	0.0000	0.0188	(0, 0.0664)	1	0
	ϕ_E	0.2000	0.1976	(0.1633, 0.2325)	0.96	0.96
III-C-a	ϕ_0	0.8000	0.7774	(0.7106, 0.8301)	0.98	0.98
	ϕ_I	0.0000	0.0198	(0, 0.0581)	1	0
	ϕ_E	0.2000	0.2028	(0.1681, 0.2381)	0.96	0.96
IV-A-a	ϕ_0	0.7000	0.6975	(0.6574, 0.7375)	0.96	0.96
	ϕ_I	0.2000	0.2041	(0.1694, 0.2397)	0.98	0.98
	ϕ_E	0.1000	0.0984	(0.073, 0.1246)	0.98	0.96
IV-B-a	ϕ_0	0.7000	0.6989	(0.6583, 0.7387)	0.94	0.94
	ϕ_I	0.2000	0.2028	(0.1679, 0.2385)	0.98	0.98
	ϕ_E	0.1000	0.0984	(0.073, 0.1246)	0.98	0.96
IV-C-a	ϕ_0	0.7000	0.6978	(0.6573, 0.7379)	0.98	0.98
	ϕ_I	0.2000	0.2004	(0.1657, 0.236)	0.98	1
	ϕ_E	0.1000	0.1018	(0.0762, 0.1285)	0.94	0.94

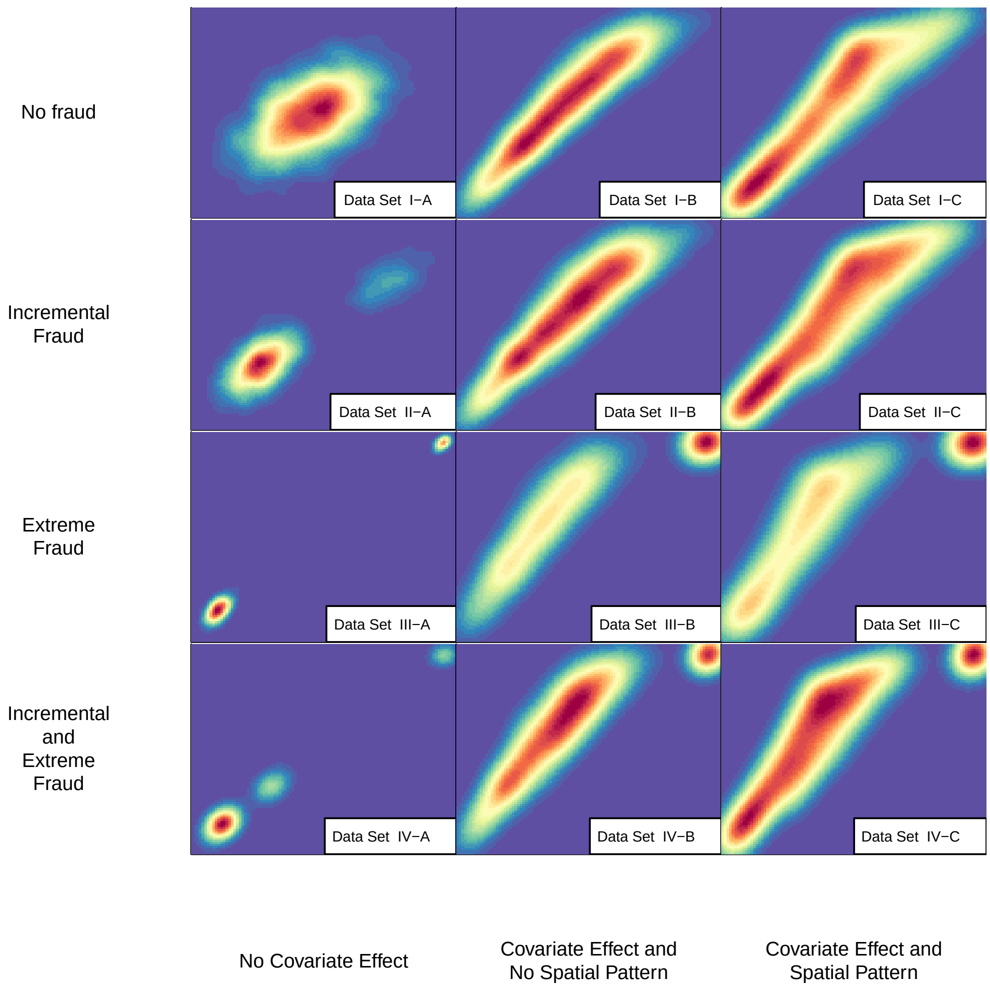

Figure 1: Contour plots of the bivariate distribution of vote proportions for the leader (w_i) and turnout $t_i = 1 - a_i$ from simluated data sets using parameters of Table 1.

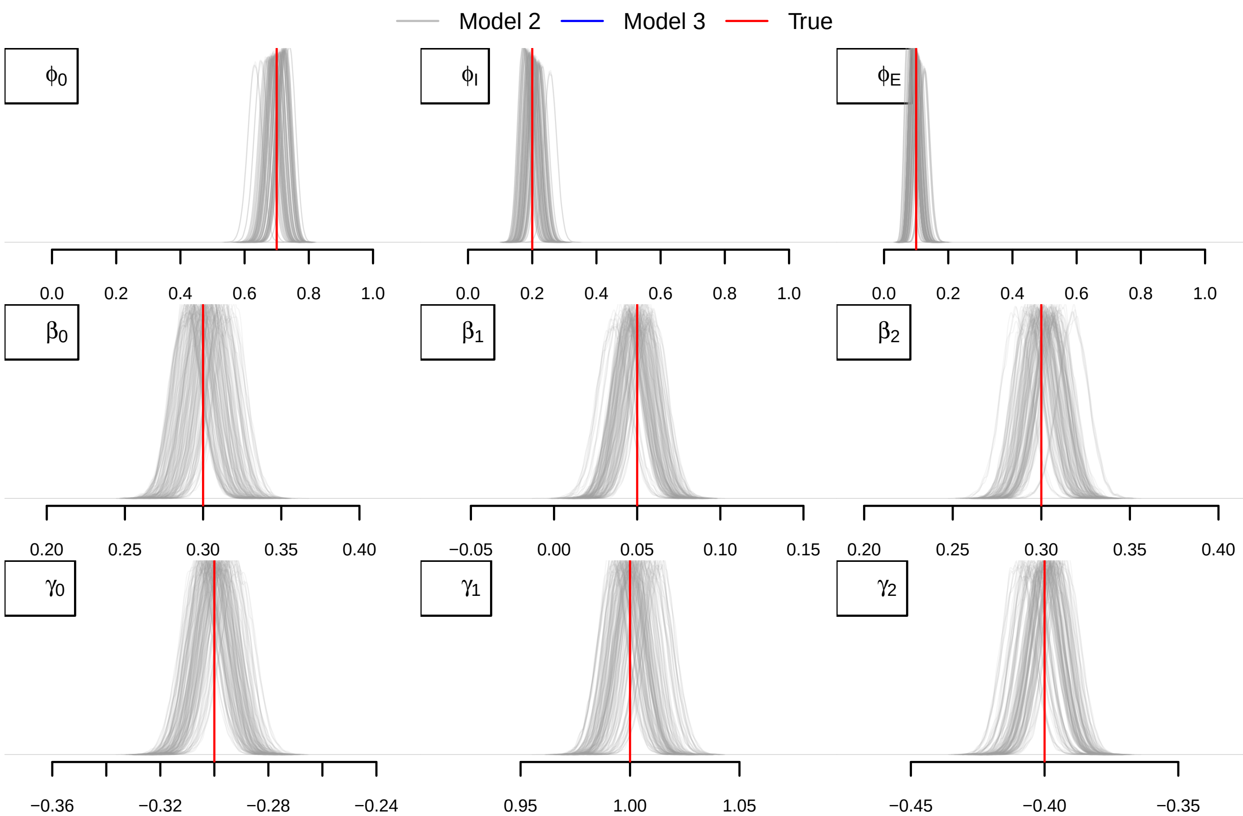

Figure 2: Posterior distribution of 100 simulated data sets using θ_{IV-B}, that is, an election with both incremental and extreme fraud, and moderate spatial concentration of the covariates.

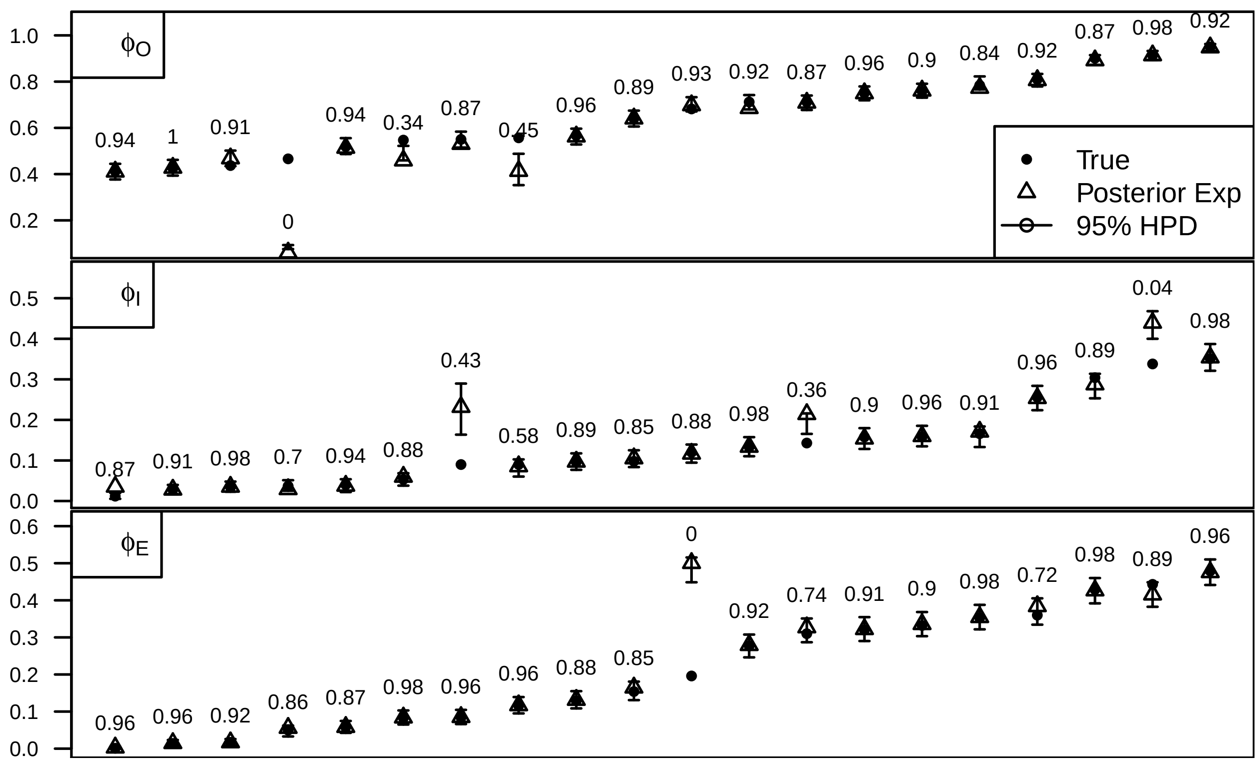

Figure 3: Estimation for each one of the 20 different parameter vector averaged across 100 data sets. Values shown in increasing order of true value for each parameter. Numbers represent the HPD coverage probability.