Positive Empirical Models of Election Fraud (that May Also Measure Voters’ Strategic Behavior)*

Walter R. Mebane, Jr.† Naoki Egami‡ Joseph Klaver§
Jonathan Wall¶

original July 22, 2014; current August 1, 2014

*Prepared for presentation at the 2014 Summer Meeting of the Political Methodology Society, University of Georgia, July 24–26, 2014. Thanks to Princeton University for supporting computing performed by Naoki Egami during his time as a visitor there during the summer of 2014.
†Professor, Department of Political Science and Department of Statistics, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: wmebane@umich.edu).
‡Undergraduate student, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902 (E-mail: naoki.egami5@gmail.com).
§Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: jrklav@umich.edu).
¶Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: jonwall@umich.edu).
Abstract

Positive Empirical Models of Election Fraud (that May Also also Measure Voters’ Strategic Behavior)

A landmark in election forensics was the 2012 PNAS paper by Klimek, Yegorov, Hanel and Thurner, which presents the first positive empirical model of election fraud: positive in the sense that the model describes what a fraudulent election looks like and then estimates the amount of fraud (of which there are two types) occurring in a particular election. The model also has the remarkable property of giving pretty much the same estimates regardless of the level of aggregation at which vote counts are observed. Being inspired by complex systems ideas, the Klimek model falls short from a statistical perspective. We modify the Klimek model to improve the shortcomings, introducing chi-squared and finite mixture variants. The resulting models do not appear to be as invariant over levels of aggregation as the original Klimek model. We show that the election fraud probability estimates from the Klimek model (using the chi-squared variant) relate meaningfully to postelection complaints in the 2009 German election and to nullification petitions in the 2006 Mexican election.¹ We also assess how well the fraud parameter estimates predict the complaints and petitions. The complaints and petitions are likely prompted in part by election frauds, so the estimated fraud parameters should relate meaningfully and predictably to them. We also show that the model also is sensitive to the strategies voters are using, at least in Germany. So what the “fraud statistics” measure may be ambiguous.

¹Practical computing difficulties prevent estimating the finite mixture variant for the Germany 2009 and Mexico 2006 data in this version of the paper.
Introduction

A landmark in election forensics is Klimek, Yegorov, Hanel and Thurner (2012), which presents the first positive empirical model of election fraud: positive in the sense that the model describes what a fraudulent election looks like and empirical in that it estimates the amount of fraud occurring in a particular election. Previously proposed methods work negatively: they attempt to detect election fraud by describing patterns “clean” elections are supposed to match and then asserting that failure to match the specified patterns indicates—or at least suggests—that election frauds have occurred (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). Mebane (2013a, 2014) argues that many of the “negative” methods, especially those based on vote counts’ second digits, are ambiguous as measures of election fraud because they also respond significantly to strategic behavior by voters as well as to other kinds of nonfraudulent political mobilization. One semipositive approach, tied to an institutional signaling argument, asserts that fraudulent vote counts should feature excesses of particular digits (Mebane and Kalinin 2009a,b; Kalinin and Mebane 2011; Shpilkin 2011; Mebane 2013b), but this approach does not imply specific estimates of the amount of election fraud.

The approach of Klimek et al. (2012) has the remarkable property of giving pretty much the same estimates regardless of the level of aggregation at which vote counts are observed. Using data at varying levels of aggregation—from the precinct level up to the regional level—Klimek et al. (2012) demonstrate this while estimating high levels of election fraud in cases that are notorious for fraudulent elections, specifically recent Russian and Ugandan elections, and negligible levels of election fraud in many other cases that are widely believed to have “clean” elections.

The methodology of Klimek et al. (2012) has some shortcomings, however. The method is inspired by complex systems ideas, but it falls short from a statistical perspective. The
parameters in the model that represent the probability of frauds—the method specifies two kinds of fraud—do not optimize a statistic that sums over all the data. The method uses data that measure both the winner’s and the opposition’s vote counts, as well as the number of registered voters, but the method’s fit is assessed using only the empirical distribution (a histogram) of the winner’s proportion of the vote. The function that summarizes the fit to the empirical distribution has no recognizable statistical properties. The procedure used to determine the parameter values has no specific properties as an estimation method. There is a plethora of random effects, seemingly as numerous as the observations. Some of the parameters that govern those effects are chosen using mechanical algorithms and others are chosen using a kind of Monte Carlo method. In the end it is impossible to say whether fraud probability values produced by the model are good estimates of the model’s parameters, and it is difficult to say how well the model as a whole describes any given data. Without being able to address such questions about the model’s adequacy or the model parameter estimates’ optimality, it is difficult to imagine any results from the model being taken all that seriously in a consequential election controversy.

Another limitation of Klimek et al. (2012) is that their argument for the efficacy of their method is merely intuitive. They find high probabilities of fraud in places that are notorious for fraudulent elections (i.e., Russia) and that have turnout and vote distributions that “look” fraudulent (i.e., bimodal with many uncompetitive places). The reliance on such intuitions may be excused because consensual measures of election fraud do not exist. Some have even cited the covert nature of election fraud as part of its definition (Lehoucq 2003). A more neutral stance may be to say that election fraud is a latent variable and one mission for election forensics is to develop ways to measure it, but here one must be careful not to conclude from the phrase “latent variable” that election fraud is in any useful sense unidimensional. Indeed, if the model in Klimek et al. (2012) describes election fraud, then election fraud is surely not unidimensional. In any case, the problem of relying on intuition to identify fraud reflects the state of the art and is
ubiquitous in the “negative” work cited above. Nevertheless it would be good to put election forensics research on firmer foundations.

We address both the statistical and the evidentiary limitations that affect Klimek et al. (2012). On the statistical front, we modify the Klimek et al. (2012) methods to assess fit using a statistically justifiable measure. This modified method still features the Klimek et al. (2012) method’s reliance on a mix of mechanical and Monte Carlo-like parameter selection methods. So we go further and frame the Klimek et al. (2012) model as a finite mixture model, which we estimate using an EM algorithm (Hasselblad 1969; Dempster, Laird and Rubin 1977). Due to a feature of the likelihood we cannot use hill-climbing algorithms in the maximization steps of the EM algorithm, so to perform the maximization we rely on GENOUD (Sekhon and Mebane 1998; Mebane and Sekhon 2011). We use a parametric bootstrap method to estimate standard errors for parameter estimates with the finite mixture specification (McLachlan and Peel 2000, 68–70).

On the evidentiary side, we use records from postelection processes in two countries that plausibly are caused in part by election frauds. From Germany we have complaints submitted after the election to the standing committee of the Bundestag that is charged with reviewing complaints about the administration of national elections. Ziblatt (2009) uses such complaints to measure the occurrence of election frauds in Germany during the years 1871–1912. Breunig and Goerres (2011, 3–4) mention that the Bundestag has never responded to a complaint in a way that “questioned the validity of an election.” We use complaints coded by type and vote data from the 2009 German Bundestag election. From Mexico we have petitions political parties submitted to the Election Tribunal of the Federal Judicial Branch (TEPJF) to request that the results of disputed casillas (ballot boxes) be nullified. In response to these petitions the TEPJF nullified the vote counts in 748 casillas (Instituto Federal Electoral 2007, 29), comprising 237,736 nullified votes in all (Tribunal Electoral del Poder Judicial de la Federación 2006). We use petition and vote data from the 2006 Mexican federal elections for President and for the House of Deputies.
We apply the several versions of the Klimek et al. (2012) model to data from the 2011 Russian Duma election. To illustrate one point we look at data from the 2012 Mexican House of Delegates elections, from the 2010 Michigan gubernatorial election and from California in the 2008 U.S. presidential election. As previously described, data from the 2009 German Bundestag election and from the 2006 Mexican presidential and deputy elections receive extensive attention.

Models

The model of elections and election frauds that Klimek et al. (2012) present is intricate but conceptually fairly simple. In the next subsection we describe their model in formal detail, but here is a less technical overview. Klimek et al. (2012) treat the number of eligible voters in each of \( n \) elections units (precincts, say) as given and fixed. \( N_i \) denotes the number of eligible voters in unit \( i = 1, \ldots, n \). The baseline assumption is that votes in an election with no fraud are produced through the interaction of processes whose effects can be summarized by two Normal distributions: there is one distribution for turnout and another, independent distribution for the proportion of votes going to the winner. Each election unit has a realized turnout proportion, denoted \( a_i \), and a realized winner’s proportion, denoted \( \nu_i \), and the total number of votes for the winner in election unit \( i \) is \( W_i = N_i a_i \nu_i \). All losing parties are treated together and considered “the opposition,” and the total number of votes for the opposition is \( O_i = N_i a_i (1 - \nu_i) \).

Klimek et al. (2012) assume that election fraud means that votes are added to the votes for the winner. Some votes are transferred to the winner from the opposition, and some are transferred from nonvoters. The 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. 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 control the magnitude of any frauds that occur.

Klimek et al. (2012) assume that election frauds do not affect the part of the
distribution of votes across election units that is to the left of certain local maximum
points. A colloquial summary of this assumption might be, “fraud occurs on the right side
of the distribution, while votes on the left are clean.” They use this assumption to estimate
the parameters that govern the Normal distributions that are assumed to describe the
distributions of nonfraudulent votes.

To estimate parameters Klimek et al. (2012) use a combination of mechanical and
Monte Carlo-like methods. They use mechanical procedures to locate the local maxima
that determine which part of the data comprise the “left” part of the data. They use
simple summary statistics to estimate the means and variance that, in their model,
describe how the nonfraudulent votes are produced, as well as one of the parameters that
describes how the election frauds operate. To estimate \( f_i \), \( f_e \) and the other parameter that
characterizes the election frauds, Klimek et al. (2012) use a protocol that involves extensive
simulation of the randomly occurring fraud shifts using pseudorandom numbers. They
select the parameter values that perform best according to a histogram-based statistic that
measures the discrepancy between the actual and simulated proportions of votes for the
winner. The protocol minimizes the discrepancy in an informal way that we describe as
Monte Carlo-like: the discrepancy is assessed using a grid of parameter values with the
simulated fraud shifts, and the parameter combinations that perform best are retained
(different pseudorandom draws are used for each set of parameter values). The results are
averaged over many iterations of the grid search.

Original Model

The Supporting Information for Klimek et al. (2012) defines the Klimek et al. (2012) model
as follows.\(^2\) The observed data for \( n \) observations indexed by \( i = 1, \ldots, n \) are the total

\(^2\)On November 5, 2013, the lead author received from Peter Klimek (via one of the lead author’s grad
students) a MATLAB script that Klimek said was used to produce the estimates in Klimek et al. (2012).
number of vote-eligible persons \((N_i)\), the number of valid votes \((V_i)\), and the number of votes for the winning party or candidate \((W_i)\). Let \(a_i = V_i / N_i\) (turnout) and \(\nu_i = W_i / V_i\) (winner’s vote proportion). Define \(\nu\) as “the number of votes where the empirical distribution function [of winner proportions \(\nu_i\)] assumes its (first local) maximum,” and define \(a\) as “the turnout where the empirical distribution function of turnouts \(a_i\) takes its (first local) maximum” (Klimek et al. 2012, SI, 1), compute the left-sided and right-sided mean deviations

\[
\sigma^L_\nu = \sqrt{\frac{\sum_{\nu_i < \nu} (\nu_i - \nu)^2}{\#\{\nu_i : \nu_i < \nu\}}} \\
\sigma^R_\nu = \sqrt{\frac{\sum_{\nu_i > \nu} (\nu_i - \nu)^2}{\#\{\nu_i : \nu_i > \nu\}}} 
\]

(1a)

(1b)

and the left-sided mean deviation

\[
\sigma_a = \sqrt{\frac{\sum_{(a_i < a) \land (\nu_i < \nu)} (a_i - a)^2}{\#\{(a_i, \nu_i) : (a_i < a) \land (\nu_i < \nu)\}}} 
\]

(2)

Define \(\sigma_x = 0.075\). Using \(\mathcal{N}(\mu, \sigma)\) to denote a normally distributed random variable with mean \(\mu\) and standard deviation \(\sigma\), the model is described as a protocol that involves two kinds of fraud and is applied to each observation \(i\). For each \(i = 1, \ldots, n\), for some \(\alpha \geq 0\) do:

1. Sample model turnout \(a^{(m)}_i \sim \mathcal{N}(a, \sigma_a)\).

2. Sample model winning vote proportion \(\nu^{(m)}_i \sim \mathcal{N}(\nu, \sigma^L_\nu \sqrt{2})\).

We used the data posted at the URL named in the paper (http://www.complex-systems.meduniwien.ac.at/elections/election.html) to verify that the code reproduces the published estimates. The code, however, varies in a number of details from the protocol described in Klimek et al. (2012), which protocol we also describe below. To produce the estimates in this paper that we describe as using the “original” model, we use either the MATLAB code written by Peter Klimek or a vectorized version we wrote in R (R Development Core Team 2011) that produces results identical to those from the MATLAB code except the R code runs about 50 times faster. We also have a version of the R code that implements the protocol described here and in Klimek et al. (2012), but results from using that code are not included in the current paper.
3. (Incremental fraud) With probability $f_i$ sample $x_i \sim |\mathcal{N}(0, \sigma^R_\nu)|$ subject to $0 < x_i < 1$ and set the number of votes for the winner as

$$W_i^{(m)} = N_i \left( a_i^{(m)} \nu_i^{(m)} + x_i \left( 1 - a_i^{(m)} \right) + x_i^\alpha \left( 1 - \nu_i^{(m)} \right) a_i^{(m)} \right),$$

(3)

the number of votes for the opposition as $O_i^{(m)} = N_i (1 - x_i^\alpha) \left( 1 - \nu_i^{(m)} \right) a_i^{(m)}$ and the number of nonvoters as $A_i^{(m)} = N_i (1 - x_i) \left( 1 - a_i^{(m)} \right)$.

4. (Extreme fraud) With probability $f_e$ sample $y_i \sim |\mathcal{N}(0, \sigma_x)|$ subject to $0 < y_i < 1$ and set the number of votes for the winner as

$$W_i^{(m)} = N_i \left( a_i^{(m)} \nu_i^{(m)} + y_i \left( 1 - a_i^{(m)} \right) + y_i^\alpha \left( 1 - \nu_i^{(m)} \right) a_i^{(m)} \right),$$

(4)

the number of votes for the opposition as $O_i^{(m)} = N_i (1 - y_i^\alpha) \left( 1 - \nu_i^{(m)} \right) a_i^{(m)}$ and the number of nonvoters as $A_i^{(m)} = N_i (1 - y_i) \left( 1 - a_i^{(m)} \right)$.

5. If neither incremental fraud nor extreme fraud is triggered, then the number of votes for the winner is $W_i^{(m)} = N_i a_i^{(m)} \nu_i^{(m)}$, the number of votes for the opposition is $O_i^{(m)} = N_i a_i^{(m)} \left( 1 - \nu_i^{(m)} \right)$, and the number of nonvoters is $A_i^{(m)} = N_i \left( 1 - a_i^{(m)} \right)$.

The intuition is that, depending on the value of $\alpha$, incremental fraud involves shifting to the winner some of the votes from the opposition and from nonvoters, while extreme fraud involves shifting to the winner almost all of those votes. Smaller values of $\alpha$ mean that larger fractions of votes are shifted from opposition to winner.

To estimate $f_i$, $f_e$ and $\alpha$ Klimek et al. (2012) average over repetitions of a grid search using the empirical distribution function of $\nu_i$, denoted $pdf(\nu_i)$.

\[^3\text{Klimek et al. (2012, SI, 1) says “}y_i \sim 1 - |\mathcal{N}(1, \sigma_x)|\text{,” but that produces values of } y \text{ near 0.0 not 1.0 and so hardly matches “extreme fraud.” The MATLAB code from Klimek (recall footnote 2) uses } y_i \sim 1 - |\mathcal{N}(0, \sigma_x)|. \text{ We credit this discrepancy to a typo in the paper.}\]

\[^4\text{To define } pdf(\nu_i) \text{ over all } i = 1, \ldots, n \text{ Klimek et al. (2012, SI, 1) write, “the data are binned with one bin corresponding to 1%.”}\]
values, where \( f_i, f_e \in \{0, 0.01, 0.02 \ldots 1\} \), \( \alpha \in \{0, 0.1 \ldots 5\} \)” (Klimek et al. 2012, SI, 1). For each set of parameters \((f_i, f_e, \alpha)\) they use the realized values \( a_i^{(m)}, \nu_i^{(m)}, x_i, y_i, W_i^{(m)} \) and \( \nu_i^{(m)} = W_i^{(m)}/V_i \) to compute the fit statistic

\[
S(f_i, f_e, \alpha) = \sum_{i=1}^{n} \left( \frac{pdf(\nu_i) - pdf(\nu_i^{(m)})}{pdf(\nu_i)} \right)^2 . 
\]

(5)

For each iteration of the grid search, they retain the set of \((f_i, f_e, \alpha)\) values that minimizes \( S(f_i, f_e, \alpha) \). To estimate \((f_i, f_e, \alpha)\) they average \((f_i, f_e, \alpha)\) over 100 iterations of the grid search procedure.

A statistical critique of the Klimek et al. (2012) model has at least three major points: their approach to assessing fit does not adequately engage the data; the functional form of the statistic \( S(f_i, f_e, \alpha) \) is strange; and their estimation procedure is at best heuristic and suggestive.

The first point is that using the empirical distribution \( pdf(\nu_i) \) is ignoring most of the data. In light of the model-defined variables \( W_i^{(m)}, O_i^{(m)} \) and \( A_i^{(m)} \), the observed data should be considered to be \( W_i, O_i = V_i - W_i \) and \( A_i = N_i - V_i \). It is not sufficient to look only at functions of \( \nu_i = W_i/V_i \) to assess how well the model fits the data. Instead, at least to keep to the nonlikelihood approach of the Klimek et al. (2012) paper, one should consider all of the discrepancies \( W_i - W_i^{(m)}, O_i - O_i^{(m)} \) and \( A_i - A_i^{(m)} \).

The strangeness of the functional form of \( S(f_i, f_e, \alpha) \) becomes clear once one does the natural thing and uses a Pearson chi-squared statistic to measure the discrepancies \( W_i - W_i^{(m)}, O_i - O_i^{(m)} \) and \( A_i - A_i^{(m)} \). That statistic is

\[
X^2 = \sum_{i=1}^{n} \left( \frac{(W_i - W_i^{(m)})^2}{W_i^{(m)}} + \frac{(O_i - O_i^{(m)})^2}{O_i^{(m)}} + \frac{(A_i - A_i^{(m)})^2}{A_i^{(m)}} \right) . 
\]

(6)

An oddity of \( S(f_i, f_e, \alpha) \) is that the squared observed empirical distribution values and not the expected empirical distribution values are in the denominator. \( S(f_i, f_e, \alpha) \) also applies
to a function (a histogram) of observed and expected proportions and not to the observed
and expected counts, hence it does not weight a discrepancy by the number of observations
from which it derives. Unlike $X^2$, whose statistical properties are well known, in statistical
terms $S(f_i, f_e, \alpha)$ is a mystery.

The estimation procedure fails to provide any confidence that the set of $(f_i, f_e, \alpha)$
values that minimize $S(f_i, f_e, \alpha)$ at each iteration are actually the minimizing set of
parameters or even that they are close to the minimizing set of parameters. It is not
reasonable to think of $S(f_i, f_e, \alpha)$ as in any sense globally concave in its arguments, so it
likely has multiple local optima. It may be reasonable to believe that each iteration
produces a result in some sense close to one of the local optima but not that that optimum
is the same across iterations. Using a grid that considers only $f_i, f_e \in \{0, 0.01, 0.02, \ldots, 1\}$
and $\alpha \in \{0, 0.1, \ldots, 5\}$ seems to reduce the chances that an iteration’s result is close to an
optimum. The results of such a Monte Carlo-like approach are at best heuristically
plausible.

**Model based on the Chi-squared Statistic**

An alternative model we consider is aimed at increasing how well the model of Klimek
et al. (2012) connects to the data. This model is essentially the same as the model and
protocol described in the previous subsection, except that we use $X^2$ from (6) to assess fit
instead of using $S(f_i, f_e, \alpha)$.

**Finite Mixture Model**

The wide assortment of random elements in the model of Klimek et al. (2012) can be a bit
confusing, but things become clearer if we change our perspective and think about the
model from a likelihood point of view. The “no fraud,” incremental fraud and extreme
fraud cases can then be seen to define three distinct components that can fit together in a
finite mixture model. Let $W, O, A$ and $N$ be vectors containing, respectively, the $n$
observations of $W_i$, $O_i$, $A_i$ and $N_i$. Because the model of Klimek et al. (2012) conditions on the $N_i$ and because $W_i^{(m)}$, $O_i^{(m)}$, $A_i^{(m)}$ are all functions of all the parameters,\(^5\) it is sufficient to express the likelihood in terms of densities for any pair of $W$, $O$, $A$. We use $W$ and $A$.

The likelihood for a finite mixture model can be written as

$$\mathcal{F}(W, A \mid N; \Psi) = \sum_{j \in \{0, i, e\}} p_j \prod_{i=1}^{n} g_{jW}(W_i \mid N; \Psi)g_{jA}(A_i \mid N; \Psi), \quad (7)$$

where $p_0$, $p_i$ and $p_e$ are probabilities with $p_0 + p_i + p_e = 1$. To adapt the language of Klimek et al. (2012), $p_0$ is the probability of “no fraud.” $g_{jW}(W_i \mid N; \Psi)$ and $g_{jA}(A_i \mid N; \Psi)$ are conditional densities and (using scalar parameters that will be defined momentarily) $\Psi = (\alpha, \nu, a, \sigma, \sigma_{\nu}, \theta)'$. The domains of the parameters in $\Psi$ are defined below.

With no fraud the densities $g_{0W}(W_i \mid N; \Psi)$ and $g_{0A}(A_i \mid N; \Psi)$ can be derived directly from the definitions Klimek et al. (2012) provide. Suppressing superscripts in the Klimek et al. (2012) model expressions, with no fraud we have $W_i = N_i a_i \nu_i$ and $A_i = N_i (1 - a_i)$. Let $\phi(x, \mu, \sigma)$ denote the Normal density with mean $\mu$ and standard deviation $\sigma$ evaluated at quantile $x$. Using $a_i \sim \mathcal{N}(a, \sigma_a)$ and $\nu_i \sim \mathcal{N}(\nu, \sigma_{\nu} \sqrt{2})$, as assumed by Klimek et al. (2012), the density of $W_i$ is

$$g_{0W}(W_i \mid N; \Psi) = \frac{\int_0^1 (\phi(W_i/\nu_i, N_i a_i, N_i \sigma_a) \phi(\nu_i, \nu, \sigma_{\nu})/\nu_i) \, d\nu_i}{\int_{-\infty}^{1} \phi(1, \nu, \sigma_{\nu}) - \int_{-\infty}^{0} \phi(0, \nu, \sigma_{\nu})} \quad (8)$$

where we use a Mellin convolution (Epstein 1948, 371) to express the density of $a_i \nu_i$, $\sigma_{\nu} \equiv \sigma_{\nu} \sqrt{2}$, and the denominator $\left( \int_{-\infty}^{1} \phi(1, a, \sigma_a) - \int_{-\infty}^{0} \phi(0, a, \sigma_a) \right)$ adjusts for the restriction $0 \leq a_i \leq 1$. The density of $A_i$ is

$$g_{0A}(A_i \mid N; \Psi) = \phi(A_i, N_i (1 - a), N_i \sigma_a). \quad (9)$$

We derive densities for $W_i$ and $A_i$ given the two fraud conditions in a similar manner,

\(^5\)Some parameters are identifiable only when the probabilities of fraud are nonzero.
by applying the stochastic assumptions of Klimek et al. (2012) in a direct way. Suppressing
the superscripts in (3) and rearranging we can write the number of votes for the winner
when there is incremental fraud as $W_i = N_i (\nu_i a_i (1 - x_i^\alpha) + x_i (1 - a_i) + x_i^\alpha a_i)$. Using
$x_i \sim |\mathcal{N}(0, \sigma_i^\nu)|$, as assumed by Klimek et al. (2012), and letting
$v(x, \sigma) = 2 \exp(-x^2/2\sigma^2)/(\sigma\sqrt{2\pi})$ denote the density of $x \sim |\mathcal{N}(0, \sigma)|$, the mean of $W_i$
given incremental fraud might be written

$$E[W_i] = N_i \int_0^1 \int_0^1 \frac{(\nu a_i (1 - x_i^\alpha) + x_i (1 - a_i) + x_i^\alpha a_i) v(x_i, \sigma_i^R) \phi(a_i, a, \sigma_a) dx_i da_i}{\text{erf} \left( 1/\sqrt{2\sigma_i^\nu} \right) \left( \int_1^1 \phi(1, a, \sigma_a) - \int_0^0 \phi(0, a, \sigma_a) \right)};$$

where $\text{erf} \left( 1/\sqrt{2\sigma_i^\nu} \right) = 2 \int_{\infty}^{1/\sqrt{2\sigma_i^\nu}} d\phi(x, 0, 1) - 1$ in the denominator adjusts for the
restriction $0 < x_i < 1$. A density for $W_i$ given incremental fraud is

$$g_{iW}(W_i \mid \mathbf{N}; \Psi) = \frac{\int_0^1 \int_0^1 \phi(W_i, \mu_i, \sigma_i) v(x_i, \theta) \phi(a_i, a, \sigma_a) dx_i da_i}{\text{erf} \left( 1/\sqrt{2\theta} \right) \left( \int_1^1 \phi(1, a, \sigma_a) - \int_0^0 \phi(0, a, \sigma_a) \right)};$$

(10a)

$$\mu_i = N_i (\nu a_i (1 - x_i^\alpha) + x_i (1 - a_i) + x_i^\alpha a_i)$$

(10b)

$$\sigma_i = N_i \sigma_i^\nu a_i (1 - x_i^\alpha) ,$$

(10c)

where we use $\theta \equiv \sigma_i^\nu$. Similarly because incremental fraud implies $A_i = N_i (1 - x_i)(1 - a_i)$,
the density of $A_i$ is

$$g_{iA}(A_i \mid \mathbf{N}; \Psi) = \frac{\int_0^1 \phi((N_i - A_i)a_i, (N_i - A_i)(1 - a), (N_i - A_i)\sigma_a) v(A_i, N_i\theta) da_i}{\int_1^1 \phi(1, a, \sigma_a) - \int_0^0 \phi(0, a, \sigma_a)}.$$

(11)

With extreme fraud we have densities for $W_i$ and $A_i$ that are derived in a similar
fashion:

\[
g_{eW}(W_i \mid N; \Psi) = \frac{\int_0^1 \int_0^1 \phi(W_i, \mu_i, \sigma_i) v(y_i, .075) \phi(a_i, a, \sigma_a) dy_i \, da_i}{\text{erf} \left( \frac{1}{\sqrt{2\theta}} \right) \left( \int_{-\infty}^1 \phi(1, a, \sigma_a) - \int_{-\infty}^0 \phi(0, a, \sigma_a) \right)} \tag{12a}
\]

\[
\mu_i = N_i (\nu a_i (1 - (1 - y_i)^{\alpha}) + (1 - y_i)(1 - a_i) + (1 - y_i)^{\alpha} a_i) \tag{12b}
\]

\[
\sigma_i = N_i \sigma_v a_i (1 - (1 - y_i)^{\alpha}) \tag{12c}
\]

and

\[
g_{eA}(A_i \mid N; \Psi) = \frac{\int_0^1 (v(A_i/a_i, 0.075(N_i)) \phi(a_i, 1 - a, \sigma_a)/a_i) \, da_i}{\int_{-\infty}^1 \phi(1, a, \sigma_a) - \int_{-\infty}^0 \phi(0, a, \sigma_a)} \tag{13}
\]

which again uses a Mellin convolution.

The finite mixture specification given by equations (7)–(13) is not quite enough to represent the idea Klimek et al. (2012) have about election fraud. Key in their distinction between \(\sigma^L_v\) and \(\sigma^R_v\) and in their left-sided definition of \(\sigma_a\) in (2) is the idea we colloquially summarize as “fraud occurs on the right side of the distribution, while votes on the left are clean.” Klimek et al. (2012) use mechanical procedures to find the “first local” maxima relative to which their \(\sigma^L_v, \sigma^R_v\) and \(\sigma_a\) are defined. We instead take a slightly cruder idea of the “left” part of the data seriously, and specify that part to be the set of \(W_i/V_i\) and \(V_i/N_i\) values that are less than each variable’s respective median. On that basis we specify upper
bounds for our parameters $\nu$, $a$, $\sigma_\nu$ and $\sigma_a$.

\[
\begin{align*}
\nu & \leq \text{median}(W_i/V_i) \quad (14a) \\
a & \leq \text{median}(V_i/N_i) \quad (14b) \\
\sigma_\nu & \leq 2 \left( \frac{1}{\#\mathcal{W}_u} \sum_{W_i/V_i \in \mathcal{W}_u} \left( W_i/V_i - \frac{\sum_{W_i/V_i \in \mathcal{W}_u} W_i/V_i}{\#\mathcal{W}_u} \right)^2 \right)^{1/2}, \quad (14c) \\
\sigma_a & \leq 2 \left( \frac{1}{\#\mathcal{V}_u} \sum_{V_i/N_i \in \mathcal{V}_u} \left( V_i/N_i - \frac{\sum_{V_i/N_i \in \mathcal{V}_u} V_i/N_i}{\#\mathcal{V}_u} \right)^2 \right)^{1/2}, \quad (14d) \\
\mathcal{W}_u & = \{ W_i/V_i : W_i/V_i \leq \text{median}(W_i/V_i) \} \\
\mathcal{V}_u & = \{ V_i/N_i : V_i/N_i \leq \text{median}(V_i/N_i) \}.
\end{align*}
\]

In the “nonfraud” case, the distributions of the winner’s vote and of turnout should not have means or variances larger than would be estimated by considering only the “left” part of their observed distributions.

Because of the behavior of $\nu(\cdot, 0, 1)$, we also require $\theta \leq \sqrt{2/\pi}$.

To use maximum likelihood to estimate all the parameters in $\Psi$ simultaneously, along with $f_0$, $f_i$ and $f_e$, we use an EM algorithm (Dempster, Laird and Rubin 1977). Because of the bounds specified for some parameters, it is possible that convergence problems such as Wu (1983) discusses—due to having parameters on a boundary—will arise.\(^7\) If fact, in practice, it is not rare to have one or more parameters reach an upper bound. Another potential problem is the identifiability problem often referred to as “label switching” (McLachlan and Peel 2000, 26–28): the different components of the mixture cannot be distinguished. This circumstance can arise if $\alpha$ is very large; then the incremental and extreme fraud distributions are similar. If $\theta$ is close to zero, the “no fraud” and incremental

\(^6\)In principle each parameter has a lower bound of zero, but in practice for computing efficiency we use positive lower bound values.

\(^7\)The distribution described by equations (7)–(13) is also not from the exponential family, which prevents a number of simplifications to the EM algorithm.
fraud components are hard to discriminate. Another kind of identification problem arises when the probabilities of fraud are zero. If \( f_i = 0 \) then \( \theta \) is not identified, and if \( f_i = f_e = 0 \) then \( \theta \) and \( \alpha \) are not identified. We ignore the affected parameters’ values once, in the EM algorithm’s iterations, either \( f_i \) or both \( f_i \) and \( f_e \) fall below a threshold.\(^8\)

A practical problem arises because the likelihood includes \( x_i^\alpha \) for \( x_i \sim |\mathcal{N}(0, \theta)| \) and \( y_i^\alpha \) for \( y_i \sim 1 - |\mathcal{N}(0, \sigma_x)| \). The half-normal distributions described by \( |\mathcal{N}(0, \theta)| \) and \( |\mathcal{N}(1, \sigma_x)| \) are generalized gamma distributions, and using \( x_i^\alpha \) and \( y_i^\alpha \) effectively presents the problem of estimating the generalized gamma distribution’s exponent (Johnson, Kotz and Balakrishnan 1994, 388). Newton-Raphson “does not work well” when used with maximum likelihood estimation to solve that problem (Johnson, Kotz and Balakrishnan 1994, 392), and other hill-climbing approaches should not be expected to perform better. We should not expect that truncating the distributions \( 0 < x_i < 1 \) and \( 0 < y_i < 1 \) and embedding them in convolutions, as we are doing, facilitates estimation. Indeed, in our experience hill-climbing optimization methods such as BFGS typically fail with this likelihood. We use GENOUD (Mebane and Sekhon 2011) to execute the maximization steps in the EM algorithm.\(^9\) The \texttt{genoud()} function effectively finds maximum points, but not being able to use derivative information makes the EM algorithm slow to run.\(^10\)

Because we cannot use derivative information and because parameter estimates are frequently on a boundary, we use a parametric bootstrap to get standard error estimates (McLachlan and Peel 2000, 68–70).

\(^8\)The threshold we currently use is \( 10^{-9} \). Once either \( f_i \), \( f_e \) fall below this threshold, the below-threshold probability is set to zero.

\(^9\)When using \texttt{genoud()} we suppress all use of BFGS. Even when applied only when the algorithm is very close to the solution, BFGS almost always causes the algorithm to wander off to undesirable and even infeasible values.

\(^10\)In most of our work so far, we have used \texttt{popsize=1000}. Trials suggest that larger \texttt{genoud()} populations may materially change the performance of the algorithm. We will explore this further in the future.
Data

To illustrate features of the three varieties of Klimek et al. (2012)’s model we have considered, we first examine data from Russia, specifically the Duma election of 2011. Unlike the analysis reported by Klimek et al. (2012) using their original algorithm, the fraud probability estimates are smaller and somewhat more variable when using either the chi-squared or the finite mixture approaches. Examples using votes from the 2012 Mexican House of Delegates elections, from the Michigan 2010 gubernatorial election and from California in the 2008 U.S. presidential election suggest the original and chi-squared versions sometimes produce misleadingly high fraud probability estimates.

Next we expand the kinds of data we use to evaluate the Klimek et al. (2012) models’ results. We use precinct-level data to estimate the models separately for each district in the 2009 German Bundestag election. We use votes from the plurality-rule “Erststimme.” We examine how the fraud probability estimates relate to the occurrence of complaints submitted after the election to the Bundestag’s Committee for Election Verification, Immunity and Rules of Procedure. We also use precinct-level data to estimate the models in each district for the 2006 Mexican federal elections for the House of Deputies and for President. We examine how the fraud probability estimates relate to the occurrence of postelection petitions submitted by the political parties to the TEPJF requesting the nullification of disputed casillas.

The general idea in the complaints and nullification petitions analyses in Germany and Mexico is that election fraud in particular localities to some extent caused the complaints and petitions to be filed regarding those localities, so if the fraud probabilities in the Klimek et al. (2012) models do measure election fraud, they should relate in a meaningful way to the chances that complaints or petitions were filed. The petitions are much better localized than the complaints are, and the subjects addressed by many of the complaints

---

11 Recall note 2.
12 Due to computing difficulties, not all the methods are used with all the types of elections in either Germany or Mexico.
seem unlikely to relate to events that would cause distortions in vote counts. Also the types of fraud represented by the model of Klimek et al. (2012) are not necessarily the only types of election fraud that may exist. Nonetheless we examine both whether the relationship between the fraud probabilities and the complaints/petitions probabilities seem meaningful and whether the fraud probabilities are effective predictors of the occurrence of the complaints and petitions. The more the relationships are meaningful and the fraud probabilities are good predictors, the more confidence we can have that the Klimek et al. (2012) model estimates are genuine measures of fraud.

One complication we consider relates to Mebane (2013a)’s finding that vote counts’ second digits relate to measures of voters’ strategic behavior, notwithstanding the claims by some (e.g. Pericchi and Torres 2011) that they can be used to diagnose election fraud. We consider evidence from the 2009 German election that suggests that at least one version of Klimek et al. (2012)’s model (the chi-squared version) also responds to measures of voter’s strategic behavior.

**Estimating the Models with Data from Russia and Elsewhere**

We estimate the original, the chi-squared and the finite mixture versions of the Klimek et al. (2012) model using data from several elections and at several levels of aggregation. Table 1 presents estimates using all three models for the Russia 2011 Duma election, for the Mexico 2012 House of Deputies election, for the Michigan 2010 gubernatorial election and for the votes from California in the 2008 United States presidential election. The Duma election uses proportional representation, while all the other elections use a plurality rule in single-member districts (or in a single district). The winner in the Duma election is the United Russia party, in Mexico the winning party varies across districts, in Michigan

\footnote{The EM algorithm we use with the finite mixture model also provides estimates of the conditional probabilities that each observation belongs to each component of the mixture (Dempster, Laird and Rubin 1977, 16): i.e., has no fraud, has incremental fraud or has extreme fraud. These quantities should be useful for determining the locations where any election frauds are likely to have occurred. We do not illustrate use of these quantities in this paper.}
the winner is the Republican party candidate and in California the winner is the Democratic party candidate.

*** Table 1 about here ***

The estimated frequency and magnitude of election frauds vary considerably across models and across levels of aggregation. The original model usually estimates higher probabilities of incremental fraud than the other models do. For Russia with the original model the estimates of $f_i$ are similarly extremely large for the territory and precinct levels of aggregation, but $f_i$ is much smaller when region data are used. This result differs from what was reported in Klimek et al. (2012) because we are using slightly different values for a threshold parameter that needs to be set to control where the original model protocol locates the “first local” maxima in the histograms.\(^\text{14}\) With the original model $f_e$ for the Russian region data is about ten times as large as with the territory or precinct data. With the chi-squared model $f_i$ is about as large for the Russian region data as for the precinct data, while $f_i$ is smaller for the territory data. The chi-squared model also produces $f_e$ for the Russian region data that is about ten times as large as the precinct data based estimate while giving an estimate of $f_e = 0$ for the territory data. The finite mixture model estimates $f_e$ to be positive for all three Russian data sets, though smaller with the territories data, while estimating a positive but small $f_i$ value for the region data, a much larger $f_i$ value for the territory data and a $f_i$ value that is again about twice as large for the precinct data. Intuitively it is implausible that the probabilities of election fraud should be the same with data at different levels of aggregation. After all we have 84 regions but 2,718 territories and 95,168 precincts. In Russia distinct fraudulent activities can occur at the territory and precinct levels. A distinctive and special distribution of election fraud across particular regions and territories would be necessary in order for the probability of

\(^{14}\)See note 2. If the same value that Klimek used (for \texttt{ElectionFitter()} argument \texttt{thres}) in his MATLAB code is used, the results are the same as were reported in Klimek et al. (2012). A point here is that there is no justified standard for how to set this argument, but the answer can strongly depend on how the argument is set.
incremental fraud to be about $f_i = .65$ for all aggregation levels, as Klimek et al. (2012) estimate. That is possible but seems unlikely to be the typical situation.

Across different levels of aggregation used with the Russia 2011 data different election fraud appears to involve distinctive patterns of vote shifts. As Figure 1 shows, finite mixture model estimates with either region data, territory data or precinct data imply that extreme fraud means that similar proportions of votes are shifted to the winner from nonvoters. Incremental fraud means that higher proportions of votes are shifted from nonvoters with region or territory data than with precinct data. The proportion of votes shifted from opposition is similar with extreme fraud and region or territory data but much higher with extreme fraud and territory data.\footnote{The point estimate for $\theta$ with territory data is $\theta = .7761$. With precinct data it is $\theta = .2697$.} Essentially no votes are shifted to the winner from the opposition through incremental fraud with region or precinct data but much high proportions are shifted with territory data.

*** Figure 1 about here ***

For the other elections we have data at only one level of aggregation for each election, but the $f_i$ and $f_e$ values vary across models. For each election the original model produces the highest value of $f_i$, the chi-squared model produces the second highest and the finite mixture model produces the lowest. $f_e$ is always zero for the Mexico 2012 Deputies elections, and $f_e$ decreases as one moves from the original model to the chi-squared model to the finite mixture model, being zero in the latter model. The finite mixture model estimates that there is no fraud at all in the Mexico 2012 Deputies or Michigan 2010 gubernatorial elections. For California $f_e$ is zero with the original and chi-squared models but slightly positive with the finite mixture model. Based on the original and chi-squared models, California 2008 has as much or more election fraud than occurred in the Russian Duma election. Intuitively that seems wrong, and on the basis of intuition the finite mixture estimates are much more credible.
The estimated magnitude of election fraud varies considerably across models. We refer to \( \alpha \). Because of the functional forms \( x_i^\alpha \) and \( y_i^\alpha \) in the Klimek et al. (2012) model, a smaller \( \alpha \) value means that larger proportions of votes are shifted to the winner when frauds occur than when \( \alpha \) is larger. It is remarkable that the chi-squared model estimates \( \alpha \) values many times larger than are estimated with the original model. For Russian precincts, for example, the original model estimates \( \alpha = 3.42 \) while the chi-squared model estimates \( \alpha = 26.9 \). The estimated magnitude of election fraud is much greater with the original model. The finite mixture model estimates magnitudes of fraud that are smaller than the original model but larger than the chi-squared model.

With the finite mixture model it is possible to motivate and compute standard errors (and in principle confidence intervals). The estimates are bona fide maximum likelihood estimates, even if the frequent occurrence of parameter estimates on a boundary makes it difficult to test this by using the gradient and the Hessian matrix. For instance, as Table 2 shows, with the Russia 2011 data the \( \sigma_{\nu} \) and \( \sigma_{a} \) parameter estimates equal their upper bounds, and the \( a \) parameter is very close to the boundary. A further complication with trying to use gradients or the Hessian to help assess the MLE is that the upper bounds on \( \nu \), \( a \), \( \sigma_{\nu} \) and \( \sigma_{a} \) are themselves estimates (medians and standard deviations) based on the data. Nonetheless, trusting the MLE to be a global maximum, we can use a parametric bootstrap to compute standard errors and confidence intervals for \( \alpha \), \( \nu \), \( a \), \( \sigma_{\nu} \), \( \sigma_{a} \) and \( \theta \) as well as \( f_i \) and \( f_e \). In Table 2 we report parametric bootstrap standard errors for these parameters. The parameters have estimates that are clearly distinct from zero.

16 McLachlan and Peel (2000, 68–70) describes the procedure.
17 We use 150 bootstrap samples.
18 The finite mixture model code used to produce the estimates reported in Table 2 has a slight, late-discovered bug that causes \( f_i \) and \( \nu \) to be underestimated. Using the correct code, the MLE has \( f_i = .0006094 \), \( f_e = .0001541 \), \( \alpha = 16.71 \), \( a = .5562 \), \( \nu = .4404 \), \( \sigma_{\nu} = .07352 \), \( \sigma_{a} = .05101 \) and \( \theta = .5139 \). \( \sigma_{\nu} \) and \( \sigma_{a} \) equal their upper bounds and \( a \) and \( \nu \) are slightly less than their upper bounds.
Germany and Postelection Complaints

We look at variation in the fraud probability parameter values across Germany in the 2009 Bundestag “Erststimmen” (single-member district plurality rule votes), and we consider how the variation in those values relates to variation in the occurrence of postelection complaints submitted to the “Ausschuss für Wahlprüfung, Immunität und Geschäftsordnung” (Committee for Election Verification, Immunity and Rules of Procedure, or AWIG). We also consider how the fraud probability estimates relate to measures of voters’ strategic behavior. We will see evidence that the fraud probability estimates relate to both.

It is usually believed that election fraud is rare in Germany, but rare does not mean nonexistent. The AWIG and hence the Bundestag did not overturn any election results in 2009 as a result of the complaints, although they did issue several recommendations on ways Germany’s elections could be improved. This reflects the AWIG’s standard that unless a complaint is shown to be “relevant” the complaint will be rejected. The committee has a particular definition of relevant: a complaint must be “mandatsrelevant,” i.e., it must demonstrably change the composition of the Bundestag or make this a possibility that cannot be ruled out (Ausschuss für Wahlprüfung, Immunität und Geschäftsordnung 2013). The Bundestag’s failure to act on a complaint does not necessarily imply that the complaint is unfounded. Facts motivating a complaint need not relate to election frauds in the sense of maleficent acts that distort votes. Because complaints generally come from citizens and not from political parties, they may be unlikely to result from partisan motivations. But some complaints are effectively partisan, such as those that relate to candidates’ inability to get on the ballot, or would-be parties’ frustration at not being treated as parties. Citizens may also act strategically with their complaints just as they do with their votes. So the ambiguity that would seem to be implied by the fraud probability parameters being related to both complaints and measures of strategy may reflect the complexity of the reality.
Using precinct data\textsuperscript{19} to estimate the chi-squared version of the Klimek et al. (2012) model separately in each of the 299 districts used in the Bundestag Erststimme in 2009 in Germany produces a narrow range of estimates: $0.00040 \leq f_i \leq 0.016$ and $0.00017 \leq f_e \leq 0.0087$, ranges that certainly match most observers’ intuitions that election fraud in Germany is rare. The geographic distribution of the estimates obtained for $f_i$ and $f_e$ can be seen in the district maps shown in Figures 2 and 3. The relatively high (but still quite low!) estimates of $f_i$ and of $f_e$ are widely scattered.

*** Figures 2 and 3 about here ***

Using different versions of the model produces estimates for $f_i$ and $f_e$ that, across versions, are substantially unrelated. Using the chi-squared model and the original model,\textsuperscript{20} both for precinct-level data, the product-moment correlation for $f_i$ is 0.012 and the correlation for $f_e$ is $-0.052$. The original model sometimes produces very high values of $f_i$. Ranges for $f_i$ and $f_e$ with the original model are $0.000 \leq f_i \leq 0.995$ and $0.000 \leq f_e \leq 0.005$.

Different types of postelection complaints occur with very different frequencies in connection with the 2009 election. Overall we distinguish twelve types of complaints.\textsuperscript{21} We code the complaint documents using a scheme that as much as possible follows the Election Incident Reporting System (EIRS) developed for elections in the United States (Verified Voting Foundation 2005; Hall 2005; Johnson 2005). Various modifications are necessary because of the particular features of the German election system. As Table 3 shows, System complaints occur in 56 of the 299 districts. The rarest type of complaint, occurring only twice, concerns Improper District Boundaries. In 2009 the mixed German electoral system produced a number of complaints that are not likely to be related to distortions in the votes in specific polling places. Indeed, System complaints, which are the most

\textsuperscript{19}Precinct data here refers to Urnenwahl and Briefwahl vote counts.

\textsuperscript{20}Due to computing issues, estimates based on the finite mixture model currently do not exist for these German data. August 1, 2014 update: estimates for the model described by equations (7)–(14a) now exist for the German 2009 Bundestag Erststimmen and are mentioned in notes 23 and 29.

\textsuperscript{21}We describe the data we collected regarding the postelection complaint process in more detail in the Appendix.
frequently occurring type, usually refer to aspects of the election system that concern overall representation, such as overhang mandates, the 5% threshold and the method used for turning votes into seats. In our judgment, the types of complaints that seem most likely to have some connection to fraudulent distortions in precinct vote counts are Ballot, Campaigning, Counting, Criminal, Party List, Polling Place and Statistics. Complaints of these types seem to connect to voters’ experiences when trying to vote, to the challenges some candidates face when trying to get on the ballot or to get a reasonable physical placement on the ballot, and to the mechanics of tallying the ballots. Other types of complaints seem to be either unrelated to the election or related only to general concerns about the electoral system.

*** Table 3 about here ***

Does the variation in $f_i$ and $f_e$ across districts reflect variation in the incidence of actual election frauds? When analyzing the relationship to postelection complaints, we assume that $f_i$ and $f_e$ do estimate the probabilities that each of two types of election fraud occur. We are not sure whether any of the types of postelection complaints are stimulated by election frauds, although as just outlined we suspect that some types are more likely to relate to incidents that affect vote counts than others are. We consider the relationship between the fraud parameters and the complaints in two ways: we examine whether $f_i$, $f_e$ and $\alpha$ relate to complaint occurrences in meaningful ways; and we test whether $f_i$, $f_e$ and $\alpha$ are good predictors of the districts where each type of complaint is located.

Both exercises are based on binomial regressions of the complaint occurrences on the estimated $f_i$, $f_e$ and $\alpha$ values. When estimating the regression models we include as regressors the two- and three-way interaction terms among $f_i$, $f_e$ and $\alpha$. Using this specification with both versions of the $f_i$, $f_e$ and $\alpha$ estimates facilitates comparing across model types, and we retain the fully interactive specification when assessing models’ performance as predictors of the complaints.
For the purpose of interpreting relationships, however, it is useful to winnow the set of regressors in order to produce a model with simply interpretable coefficients. For this we rely on the AIC statistic. Not only does AIC indicate that many of the two- and three-way interaction terms are not necessary, but in fact the AIC statistic along with simple $t$-tests suggests that in many cases none of the coefficients are statistically significant at all. Table 4 shows residual deviances for the complaint types that have at least one statistically significant coefficient with either the chi-squared or the original Klimek et al. (2012) models. The residual deviance is shown for the specification that has the smallest AIC statistic among the specifications with all combinations of single and interacted variables. A coefficient is deemed significant, for this purpose, if the $p$-value of the coefficient’s $t$-statistic is less than .10. Six types of complaints have at least one significant coefficient when using the chi-squared model’s $f_i$, $f_e$ and $\alpha$ estimates and two of those types have at least one significant coefficient when using the original model’s estimates. The six types are Polling Place, Counting, Criminal, Campaigning, Statistics and Party List. These six types are all among the seven types we expected to have some connection to fraudulent vote distortions (Ballot is the exception). Table 4 also shows for these complaint types the residual deviance for the null model, which includes only the Intercept term. The specifications that include coefficients for some of $f_i$, $f_e$ and $\alpha$ do not generally reduce the residual deviance by all that much.

*** Table 4 about here ***

As Table 5 shows, some of the coefficients in the specification that have the smallest AIC value are barely or not at all significant. The sole coefficients in the specifications for Counting and for Party List are marginally significant. In the specification for Criminal the only marginally significant coefficient is the coefficient for $\alpha$. For Polling Place,

---

22 The $t$-test statistics are not corrected for multiple testing.

23 With the finite mixture model estimates of $f_i$, $f_e$ and $\alpha$ the regressions for many types of complaints fail due to quasiperfect separation problems; a weighted combination of the covariates perfectly separates the districts with complaints from those without complaints. Such a results suggests the finite mixture model fraud probability estimates relate very strongly to the occurrence of many of the types of complaints.
Campaigning and Statistics, however, not only are coefficients for at least one of $f_i$, $f_e$ or $\alpha$ significant but so are the coefficients for one of the interaction terms $f_if_e$ or $f_i\alpha$. Such interactions render the coefficient estimates not directly interpretable. To try to understand what effects the fraud parameters have on the incidence of Polling Place, Campaigning and Statistics complaints, we use the estimated coefficients to simulate the probability that a complaint of each type is made. To do this in a way that reflects the actual distribution of $f_i$, $f_e$ and $\alpha$ across districts, we use the actual data to perform the simulations.

*** Table 5 about here ***

To summarize and display the simulation results, we use nonparametric regression.\footnote{We use the sm (Bowman and Azzalini 1997) package of R to compute the nonparametric regressions.} Because $f_i$, $f_e$ and $\alpha$ are all relevant and three-dimensional graphs are difficult to interpret, we show results regressed on $f_i$ and $f_e$ separately for $\alpha$ above and below the median value of $\alpha$ across the districts. Recall that a smaller $\alpha$ value means that larger proportions of votes are shifted to the winner when frauds occur than when $\alpha$ is larger. Figure 4 shows the patterns in the simulated probabilities this method produces. For Polling Place complaints (Figure 4(a,b)), the value of $\alpha$ affects the magnitude of the probabilities but not how the probabilities relate to $f_i$ and $f_e$: the probability of a Polling Place complaint tends to increase as either $f_i$ or $f_e$ increase, or as both increase. The probability of a Campaigning complaint tends to increase with $f_e$ for lower values of $f_i$ but the probability tends to decrease as $f_e$ increases for higher values of $f_i$ (Figure 4(c,d)). For Statistics complaints the direction of the relationship with the probabilities is difficult to discern when $\alpha$ is large, but when $\alpha$ is small the probability of a Statistics complaint tends to decrease as $f_e$ increases.

*** Figure 4 about here ***

For Counting and Party List complaints we can determine that the probability of a complaint tends to decrease as $f_i$ increases directly from the coefficient estimate in Table 5,
but it may be useful to see how that effect translates through the logistic functional form and to get a sense of the precision of the estimated relationship. Figure 5 shows the simulated probability graph along with 95% confidence bounds. The probability of a Party List complaint decreases for lower values of \( f_i \), as does the probability of a Counting complaint, but the latter probabilities fall to zero whenever \( f_i \) is greater than \( f_i \approx 0.003 \). The probability of a Party List complaint does not fall to zero until \( f_i \) is greater than \( f_i \approx 0.006 \).

*** Figure 5 about here ***

All of the complaint types with significant coefficients are among the types we expect to connect to fraudulent distortions in vote counts. Of the expected set, only Ballot complaints fail to have a significant coefficient.\(^{25}\) Why the probabilities of Counting and Party List complaints tend, slightly, to decrease as \( f_i \) increases is a question. On the whole, however, the relationships between the chi-squared election fraud parameters and the probabilities of postelection complaints seems reasonable.

How well can we predict the complaints based on \( f_i \), \( f_e \) and \( \alpha \)? If the estimated fraud parameters capture some aspects of electoral fraud, we may expect that not only will they have meaningful relationships with the probabilities of complaints but they will measurably add to the ability to predict where the complaints will occur. To do this we use the fully saturated model in a cross-validation exercise. We compare the original and chi-squared estimates to one another. We also compare both of these methods to a set of \( f_i \) and \( f_e \) estimates chosen by random sampling the uniform distribution.\(^{26}\) The predictive model is the logit regression model and includes main effects, all two-way interactions and the three-way interaction of \( f_i \), \( f_e \) and \( \alpha \). We compare the three versions by prediction error

\(^{25}\)With Ballot type complaints and original model \( f_i \), \( f_e \) and \( \alpha \) estimates, the regression terminates with what looks to be a case of quasiperfect separation. The best \texttt{glm(family=binomial())} (R Development Core Team 2011) model according the AIC statistic is \texttt{Ballot \sim f_i * f_e + \alpha} , but for instance \( f_i \) has a coefficient (SE) of \( 4.198e+15 \ (7.358e+07) \). the coefficients for \( f_e \) and for \( f_i * f_e \) are, respectively, \( 1.470e+17 \ (6.935e+09) \) and \( -9.034e+14 \ (1.116e+07) \). \texttt{glm()} warns, “glm.fit: fitted probabilities numerically 0 or 1 occurred.” The positive coefficient for \( f_e \) may suggest the probability increases with \( f_e \), but it’s hard to say.

\(^{26}\)The random value selection uses the maximum of the two estimates as max, and the minimum of estimates as min.
relative to the random baseline predictors.\textsuperscript{27}

The boxplots in Figure 6 show that the chi-squared model’s parameter estimates do not always predict well in cases where the coefficient estimates suggest there are meaningful relationships between the combination of $f_i$, $f_e$ and $\alpha$ and the probability of a complaint. To predict well here means that a model outperforms the random model. The chi-squared model does well for Polling Place complaints and somewhat well for Party List and Counting complaints. The chi-squared model also does well for Ballot, which we expect to relate to vote distortions but does not have significant coefficients, as well as for Allegations, which is less obviously connected to vote distortions. The chi-squared model does not predict well for Statistics, Campaigning or Criminal complaints. The original model does well for all the types of complaints, including the System type for which it is not easy to imagine a connection to distortions in the vote counts.\textsuperscript{28} As most of those prediction successes for the original model depend on meaningless models—models that are devoid of significant coefficients—it may be best to judge these successes as examples of the frequent finding that the models that predict best are often not the models that are meaningful or faithfully represent causal relations.

*** Figure 6 about here ***

We take the occasional correspondence between meaningful relationships and predictive success to be more reassuring than discouraging about the chi-squared version of the Klimek et al. (2012) model. We think the original model’s predictive success even in instance where there is little reason to expect vote-based fraud measures to relate to a type of complaints is worrisome. While the original model’s estimates are connecting to something, it seems unlikely that that thing is predominantly election fraud.

While the chi-squared version of the Klimek et al. (2012) model comes through its engagement with the postelection complaints in moderately good standing, an interpretive

\textsuperscript{27}We compute 50 replicates using the \texttt{cv.glm()} function in the \texttt{boot} package of \texttt{R} (Davison and Hinkley 1997). The prediction error is normalized based on the mean and standard error of the Random model.

\textsuperscript{28}The original model also performs well predicting ID type complaints.
complication arises when we consider whether the estimates of $f_i$, $f_e$ and $\alpha$ relate to measures of voters’ strategic behavior. The basic dilemma, as Mebane (2013a, 2014) articulates, is that votes may differ from preferences—in other words, votes may be shifted—not only due to election frauds but also because voters shift votes themselves when they act strategically. Voters in Germany are well known to behave strategically. The mixed system gives opportunity to observe different distributions of votes being cast in the same district at the same time under both plurality (Erststimme) and proportional representation (Zweitstimme) rules. The difference between those votes is often used as a measure of strategic behavior (e.g. Cox 1997, 83; Bawn 1999). With the plurality election results alone, measures such as the difference between the winner’s and the third-place candidate’s votes connect to strategic behavior (Cox 1994). The conditional mean of the second significant digits of a candidate’s votes, which Mebane (2013a, 2014) studies, relates to strategic behavior perhaps in addition to election frauds.

Figures 7–11 show there are strong relationships between the district-specific chi-squared model estimates of $f_e$ and all of the aforementioned measures of voters’ strategic behavior. The figures show the results from nonparametrically regressing $f_e$ on pairs of the measures. The figures include reports of the $p$-values for tests of the nonparametric regression model compared to the model of no effects. The effects are least significant when related to the Zweitstimmen-Erststimmen differences in the votes for CDU/CSU, but for all the other parties they are strongly significant. The means of the second digits of the second-place candidate’s votes are not importantly related to the values of $f_e$, but the second-digit means of the winning candidate are.29

*** Figures 7, 8, 9, 10 and 11 about here ***

Using the Klimek et al. (2012) model in Germany ultimately produces ambiguous results. At least the chi-squared model’s estimates of $f_i$, $f_e$ and $\alpha$ meaningfully and

29When the finite mixture model estimates of $f_i$, $f_e$ and $\alpha$ are used there are no significant relationships with the various measures of strategic behavior.
somewhat effectively relate to postelection complaints about the integrity of the electoral process. But those same estimates very strongly relate to measures of voters’ strategic behavior.\footnote{See note 29.} Perhaps the baseline very low level of election frauds in Germany—which the very small magnitudes of the chi-squared model estimates for $f_1$ and $f_e$ do capture—allows the Klimek et al. (2012) model to pick up the distortions in vote distributions that voters (who are not malefactors) cause by their strategic voting activities.

**Mexico and Nullification Petitions**

We look at variation in the fraud probability parameter values across Mexico in the 2006 federal elections for House of Deputies (single-member district plurality rule votes) and for President, and we consider how the variation in those values relates to variation in the occurrence of petitions submitted by the political parties to the TEPJF requesting the nullification of disputed presidential casillas.\footnote{Presidential casillas are physically distinct from the casillas used in the House of Deputies elections. The ballot boxes are often placed next to one another on the same table to receive the paper ballots, but they are separate boxes.} We describe the petition process and the data we collected regarding that process in more detail in the Appendix. A summary is that petitions were filed regarding 27,109 casillas.\footnote{In all, voting data show there were 130,788 casillas.} The petitions referenced one or more of eleven types of incidents, labeled A–K (brief descriptions of the types appear in Table 6, and the Appendix describes these types in detail). One reason the TEPJF nullified the vote counts of only 748 casillas is because several of the types require that the incident be “determinative for the outcome of the vote,” which, according to the court, means that the number of votes in error is greater than the difference between the candidates finishing first and second in the casilla’s votes (Ochoa 2000, 7). Table 6 shows how often each type of incident claim occurs. The most frequent type is F (“Willful misconduct or error in the vote count”), while the least frequent is H (“Impeding the ability of political party representatives to observe elections at the polling place”). We think all of these types of
claims may trace back to election frauds, although not all to the same kinds of election fraud. Of course, many may stem from accidents and some may be entirely phony claims.

*** Table 6 about here ***

Using sección (precinct) data to estimate the chi-squared version of the Klimek et al. (2012) model separately in each of the 300 districts for the House of Delegates election in 2006 in Mexico produces a wide range of estimates. The range of the estimates obtained for $f_i$ and $f_e$ can be seen in the district maps shown in Figures 12 and 13. $f_i$ is near 1.0 in a few districts, but in most districts $f_i$ is near zero or exactly zero. The maximum value of $f_e$ is 0.0105, and only a few districts have $f_e$ values nearly that high. Mostly $f_e$ is zero.

*** Figures 12 and 13 about here ***

Because we study how variation across districts in the fraud probabilities relates to the occurrence of nullification petitions, we need to estimate $f_i$ and $f_e$ separately for each district for the votes in the presidential election. This presents a problem of deciding who is the “winner” for use in the Klimek et al. (2012) model. PAN (Partido Acción Nacional) won the election on a national basis, but other parties often had the most votes in particular districts. Indeed, in one district the PAN candidate received only 2.7% of the votes. Hence we estimate the model (specifically the original version of the model) using two alternative procedures. For one alternative PAN is treated as the winner everywhere, and for the other alternative we always use the party that received the most votes in the district as the “winner.” Figure 14 shows that treating PAN as the winner everywhere and using the original Klimek et al. (2012) method produces many more high values for $f_i$ than were observed with the Deputies data and the chi-squared method. Figure 15 shows that $f_e$ with the presidential data has a higher maximum value ($\max(f_e) = 0.024$) than occurs with the deputies data, but fewer districts have $f_e$ values near the observed maximum with the presidential data than with the deputies data.

33“Everywhere” in fact means almost everywhere. In five districts the model could not be estimated. In those five cases we use the party that received the most votes in the district as the “winner.”
Using differing versions of the model or differing levels of aggregation in the analysis produces estimates for $f_i$ and $f_e$ that, across versions or levels, sometimes are fairly strongly related. Table 7 shows product-moment correlations between the fraud parameters using several different implementations: the chi-squared model estimated using deputies data at the sección level; the original model using deputies data at, respectively, the sección level and the casilla level; and the original model estimated using, respectively, the two presidential vote implementations both with sección level data. The correlations for $f_i$ are all positive, although some are very near zero. For $f_e$, interestingly, the chi-squared deputies-based estimate and original model presidential-based estimates are fairly strongly and positively correlated. The sección-level deputies-based estimate using the original model is moderately strongly correlated with the casilla-level estimate using the same model. Other correlations for $f_e$ are near zero or even negative. The moderately strong correlations for both $f_i$ and $f_e$ between the estimates using sección-level data and the estimates using the casilla-level data give some support to Klimek et al. (2012)’s claim that the level of aggregation of the vote data does not very much affect the results. The correlation between the chi-squared deputies-based estimates and the estimate based on presidential votes suggests that frauds that affect the respective offices tend to happen in the same geographic locations—or least in the same legislative districts.

Does the variation across districts faithfully reflect variation in the incidence of actual election fraud? We maintain the assumption that $f_i$ and $f_e$ do estimate the probabilities that each of two types of election fraud occur. We do not know how many of the types of nullification petitions are stimulated by election frauds, but we suspect that many of them are. As we did in the case of German postelection complaints, we consider the relationship

\footnote{Due to computing issues, estimates based on the finite mixture model currently do not exist for these Mexican data.}
between the fraud parameters and the petitions by examining whether $f_i$, $f_e$ and $\alpha$ relate to nullification incident claim occurrences in meaningful ways and by testing whether $f_i$, $f_e$ and $\alpha$ are good predictors of where each type of incident claim is filed.

Both exercises are based on overdispersed binomial regressions of the petition occurrences on the estimated $f_i$, $f_e$ and $\alpha$ values.\textsuperscript{35} The regression specification is designed to respect the fact that the nullification petitions vary at the level of individual casillas, while the fraud parameter estimates vary only at the level of legislative districts. We aggregate the petition counts to the sección level: the dependent variable is the count of how many casillas in each sección do or do not have each type of nullification incident claim.\textsuperscript{36} The regression model specifications include three covariates, computed from presidential voting data, in addition to the $f_i$, $f_e$ and $\alpha$ values.\textsuperscript{37} These values are: the margin between the leading and second-place candidate in the sección in the presidential race, as a proportion of all votes cast in the sección; the proportion of votes cast for the leading candidate in the sección in the presidential race; and the absolute difference between the PAN and PBT (coalición “Por el Bien de Todos”) parties in the sección.

We use these covariates in order to capture the following political incentives that likely affected which casillas the parties petitioned to nullify. The TEPJF requirement that the numerical error in a casilla be greater than the difference in the casilla between the candidates finishing first and second in the casilla applies to incident types F, G, I, J or K. As a result, parties in 2006 may have been more likely to challenge a particular casilla when the margin between first and second place was small. In these cases we rely on the margin variable. Nullification petitions are also the parties’ last means of altering the vote count in their favor. Because the outcome of the vote nationally was quite close (PAN won by 0.58% in the end), the parties—specifically PBT—may have been more likely to

\textsuperscript{35}We use the function \texttt{glm(family=quasibinomial())} in \texttt{R} to estimate the regression models.

\textsuperscript{36}Ideally the casilla-level variation would be preserved, but time did not permit us to complete the process of matching the codes used to refer to casillas in the TEPJF source documents to the codes used to refer to casillas in the voting data.

\textsuperscript{37}The data used to compute the covariates are the data that existed before many counts were changed through various recounts.
challenge a particular casilla when the winning margin favoring PAN was large, in the
interest of gaining back the most votes. A very large winning proportion may also in some
cases have seemed suspicious. In these cases we rely on both the winner proportion
variable and the absolute PAN-PBT margin variable.

When estimating the regression models we include the two- and three-way interaction
terms among $f_i$, $f_e$ and $\alpha$ among the regressors, while the three covariates derived from the
prerecount president vote data enter linearly. Using this specification with all the varieties
of $f_i$, $f_e$ and $\alpha$ estimates facilitates comparing across model types. Table 8 shows that
when the residual deviance is used to compare across the $f_i$, $f_e$ and $\alpha$ estimates that are
based on sección data, no one type of Klimek et al. (2012) method dominates. Every
method always has a significantly better fit when $f_i$, $f_e$ and $\alpha$ are included in the model
than when they are excluded and only the covariates are included (that is the “null”
model). Every model except the one based on the district-by-district presidential winner
has the best fit with at least one type of nullification petition.

*** Table 8 about here ***

We would like to interpret the coefficient estimates from the regressions to determine
which if any of the types of nullification incident claims tend to increase with $f_i$ or $f_e$. 
Unfortunately for this purpose, the AIC statistic in almost all cases suggests that all of the
two- and three-way interaction terms are necessary. Including so many interaction terms
renders the coefficient estimates not directly interpretable. To try to understand what
effects the fraud parameters have on the incidence of nullification petitions, we use the
estimated coefficients to simulate the probability that a nullification claim of each type is
made. To do this in a way that reflects the actual distribution of $f_i$, $f_e$ and $\alpha$ across
secciones, we use the actual data to perform the simulations, except we replace the three
covariates that were computed from the presidential vote data with those variables’ median
values.
To summarize and display the simulation results, we again use nonparametric regression. Figure 16 shows a representative selection of the displays this simulation cum nonparametric regression method produces. Figure 16(a) shows that when \( \alpha \) is “large” (above the median), the probability of a casilla having a nullification claim of type F tends to increase with \( f_e \) when \( f_i \) is large but decrease as \( f_e \) increases when \( f_i \) is small. When \( \alpha \) is large the probability of F tends to increase with \( f_i \) when \( f_e \) is relatively large but does not vary with \( f_i \) when \( f_e \) is relatively small. In contrast, Figure 16(b) shows that when \( \alpha \) is “small” (below the median), the probability of F always increases with \( f_e \). When \( \alpha \) is small, we still observe that the probability of F does not vary with \( f_i \) when \( f_e \) is relatively small. Figures 16(c–f) show the patterns with which the probabilities of nullification claims of types I and J vary depending on \( f_i \), \( f_e \) and \( \alpha \). Typically the probabilities increase as \( f_i \) or \( f_e \) increase, or as both increase, but in Figures 16(d,e) the probability decreases as \( f_e \) increases when \( f_i \) is small.

*** Figure 16 about here ***

The patterns in Figure 16 represent the patterns observed for most of the types of nullification petition claims. In terms of the way the probabilities tend to increase or decrease with \( f_i \) and \( f_e \) depending on whether \( \alpha \) is “large” or “small,” the pattern for claims of types E and K are pretty much the same as for type F, and the claims for B and G are pretty much the same as for type I. As much as the pattern for J resembles the patterns of F and I, so does the pattern for C. As can be seen in Figure 17, the patterns for A, D and H are not quite as similar. For each of these latter three types, the probability tends to increase with \( f_e \) and somewhat with \( f_i \) when \( \alpha \) is small. When \( \alpha \) is large the probability of A or D often tends to decrease as \( f_e \) increases, while the probability of H tends to increase with \( f_e \).

---

38If one looks separately at the left and right sides of each of the graphs in Figure 16, then three of those sections for J resemble corresponding sections for F and for I. Such a “three-quarters” similarity also holds for type C, as well as between F and I.
How well can we predict the nullification petition claim types based on $f_i$, $f_e$ and $\alpha$? To answer this question we again use the fully saturated model that includes the covariates in a cross-validation exercise, comparing the original and chi-squared estimates to one another. We also compare both of these methods to the set of regressors that includes only the three covariates we derive from the presidential vote data and to models that include the $f_i$, $f_e$ and $\alpha$ estimate that are based on presidential vote data. The predictive model is the overdispersed binomial regression model and includes main effects, all two-way interactions and the three-way interaction of $f_i$, $f_e$ and $\alpha$, plus the three covariates. We compare the five versions by prediction error relative to the Covariates Only baseline predictors.\footnote{We compute 50 replicates using the \texttt{cv.glm()} function in the \texttt{boot} package of \texttt{R} (Davison and Hinkley 1997). The prediction error is normalized based on the mean and standard error of the Covariates Only model.}

The $f_i$, $f_e$ and $\alpha$ estimates often do not increase the ability to predict the nullification petition claim types by a lot, but they do do so noticeably and consistently. As shown in Figure 18, the median prediction error for the parameters estimated using the chi-squared model is lower than the median prediction error in the Covariates Only model for all types of petition claims. In some cases, as for example type K, the prediction improvement is greater than in other cases. The models that use the original model’s estimates have prediction superior to the Covariates Only model somewhat less frequently. The models that use $f_i$, $f_e$ and $\alpha$ estimates based on the presidential vote data usually but not always predict claim type occurrences better than the chi-squared model does. Again K is the type that stands out for superior predictive performance by the chi-squared model.

Discussion

In terms of the model of the data it implies—in terms of the likelihood it specifies—the Klimek et al. (2012) model seems to capture important aspects of election fraud.
Maleficent distortions in the votes likely are most often distortions in favor of the winner, either a moderate or an extreme amount of votes might be misdirected, and the purloined votes can be taken either from the opposition or from nonvoters. All that is reasonable. As a default it is not crazy to assume that turnout and the proportion of votes for the winner are normally distributed across the units in which votes are observed: such would be the simple large-sample intuition about how preferences are distributed. The functional form involving exponentiation by $\alpha$ to measure the magnitude of frauds presents statistical estimation challenges but is inherently an interesting idea.

The protocol Klimek et al. (2012) propose for parameter estimation, however, leaves much to be desired. We have shown that even the basic step of using a better measure of fit to the data—using $X^2$ instead of $S(f_i, f_e, \alpha)$—produces more meaningful parameter estimates. With the data from the German 2009 election, the chi-squared method that uses $X^2$ leads to significant and (sometimes evaluated through simulations) meaningful coefficients in models that relate $f_i$, $f_e$ and $\alpha$ to postelection complaints. Klimek et al. (2012)'s original method, which uses $S(f_i, f_e, \alpha)$, leads to sometimes greater effectiveness in predicting the complaints but the predictions and the usually insignificant coefficients that support them are meaningless. With the data from the Mexican 2006 House of Deputies election, the chi-squared method does well relating to and predicting the different types of nullification petition claims. In the estimates of $f_i$, $f_e$ and $\alpha$ for the Russian 2011, Mexican 2012 and Michigan 2010 elections and for the votes in California for the 2008 U.S. presidential election, we have only intuition to rely on to interpret and evaluate the results. But even so Klimek et al. (2012)'s original method seems to produce estimates of $f_i$ and $f_e$ that are too high: for instance, California in 2008 does not have more election fraud than Russia in 2011. The chi-square method reduces the parameter estimates somewhat, but the finite mixture method produces what look like the most plausible results.

When combined with a credible estimation methodology, the Klimek et al. (2012) model seems suitable to become an effective tool for estimating the incidence and
magnitude of election frauds. Because one model parameter is the mean proportion of votes for the winner in the absence of fraud—this parameter is \( \nu \)—the Klimek et al. (2012) model may be a basis for determining the number of votes the winning and opposition sides would have received had there been no fraud. To take the estimate of \( \nu \) so seriously will depend on clarifying exactly how plausible are the basic assumptions that in the absence of fraud turnout and the winner’s proportion of the votes follow Normal distributions. While those assumptions are not crazy, it’s not obvious they are convincing. Perhaps to boost the credibility of the assumptions it will be necessary to make \( \nu \) and the turnout mean parameter \( a \) functions of covariates. The conceptual adequacy of the model will also need to be addressed. We need to decide whether the “fraud parameters” measure only election frauds or also pick up strategic voting.\(^{40}\) Other conceptual limitations, which we touch on below, also would need to be resolved.

Even when realized through a finite mixture likelihood specification, the Klimek et al. (2012) approach is not perfect. It remains to be seen whether estimates from the finite mixture model are subject to the same ambiguity as affects the chi-squared model. The chi-squared model’s estimates for \( f_i \), \( f_e \) and \( \alpha \) relate strongly to measures of strategic voting behavior.\(^{41}\) The estimated levels for \( f_i \) and \( f_e \) are extremely low in the German 2009 election, and maybe the fact that there is little or no election fraud in Germany allows the way the Klimek et al. (2012) model’s parameters respond to voters’ strategic actions to become apparent. The parameters may be said in this instance to measure strategic behavior along with election fraud. We reserve judgment on whether this is an inherent feature of the Klimek et al. (2012) model until we can analyze the estimates from the finite mixture model applied by district to the German 2009 data.

Even as useful as it may be, the Klimek et al. (2012) model has significant conceptual limitations. We mention three concerns. First, it is clear that the two types of vote

---

\(^{40}\)Recall note 29.

\(^{41}\)While not reported in the current paper, relationships between combinations of \( f_i \), \( f_e \) and \( \alpha \) and strategic voting measures are even more pronounced in data from the 2005 German Bundestag election.
manipulation that appear in the Klimek et al. (2012) model are not the only types of manipulation that exist. The most important manipulation the model is ignoring is voter suppression: some who should be or are eligible to vote are prevented from voting. Voter suppression does not add votes to any party, so it is not at all in line with the spirit of Klimek et al. (2012)’s conception. If voters who support one party are more likely to be suppressed than are voters who support the winning party, say, then it is possible the Klimek et al. (2012) model will capture the manipulated turnout and voting distributions somehow, but voter suppression will be represented only in an aliased way: \( f_i \) or \( f_e \) may be significantly positive in this case, but the form of the frauds will be misdiagnosed. A second concern is the assumption that all election frauds benefit the winning party with all opposition parties being treated as if they are a unit and all equally victims. The model ignores the possible diversity of election frauds. Our third concern is that systems with multiple winners—when the district magnitude is greater than one—are not easy to accommodate. In the current model, votes for all winning parties or candidates have to be summed to produce a single “winner’s” vote. Possibly proportional representation systems should be considered to be systems where there is more than one winner, in which case for example it is questionable to apply the Klimek et al. (2012) model to the Russia 2011 Duma election data.

Data Appendix

German complaints data

One of the standing committees of the Bundestag is the “Ausschuss für Wahlprüfung, Immunität und Geschäftsordnung” (Committee for Election Verification, Immunity and Rules of Procedure). This committee deals with the rules of the Bundestag, possible criminal proceedings against Bundestag members and complaints about the administration of national elections (Bundestagswahlen, Europawahlen, etc.).

All of the complaints data come from the archives of the Bundestag’s website. The “Drucksache” field represents the document number in the form “Election Period/Document Number.” There is also a file number associated with every complaint in the form “WP XX/Election Year.” The Drucksache field and the file number allow for the easy finding of the original complaint’s text. The name, location, and reason fields are all taken directly from the original documents published by the relevant Bundestag
committee. The “EIRS Coded Reason” borrows from the Election Incident Reporting System (Verified Voting Foundation 2005; Hall 2005; Johnson 2005), with a few additions necessitated by the vagaries of the German electoral system and the type of complaints that it precipitates. To determine the Wahlkreis (electoral district) that corresponds with the zip code given for each case, we used a shapefile of German zip codes in conjunction with a shapefile that shows the Wahlkreis boundaries for the relevant election (occasionally, these borders were unclear, most likely due to projection differences between the two shapefiles). It should be noted that the locations given in the files are only the location of the complainant, i.e., it is entirely possible for someone to complain about an issue that they themselves did not experience—standing is not an issue.

Here is a list of all of the reason codes we applied to the complaints data. They translate the nature of the complaint as presented in the long-form text into a more database-friendly form. We refer to our set of codes as EIRS+.

- Absentee-ballot related problem: cases where complainants did not receive their absentee ballot, their absentee ballot came late, or where there were any other problems related to the preparation or administration of absentee voting.
- Registration related problem: cases where complainants were not able to vote or request an absentee ballot due to problems with their registration (not registered at all or they were registered in a different location) or in cases where there were problems mailing the Wahlbenachrichtungen (letters that notify registered German voters when an upcoming election will take place and where they are supposed to vote).
- Improper Campaigning Influence: cases where the complainant encountered improper campaign advertising (for example, advertising too close to a polling place) or felt that any of the parties’ campaigns were conducted in an otherwise inappropriate, if not necessarily illegal, manner.
- ID related problem: as Germany does not have strict voter ID laws, many complainants demanded a more robust process for checking the identity of voters at the polling place.
- Criminal status related problem: cases where problems with the administration of federal elections in prisons were alleged.
- Disability access problem: cases where polling places or other voting-related buildings were not accessible to the disabled.
- Ballot related problem: all complaints related to the physical characteristics of the ballot and its design (size of ballot, color of ballot, order of candidates on ballot, etc.)
- Polling place problem: includes problems related to the built environment of the polling place, the set-up of the voting booths and other temporary election structures, as well as problems with polling place workers. A few examples of

---

42See [http://arnulf.us/PLZ](http://arnulf.us/PLZ).
problems with the built environment of a polling place would be the presence of surveillance cameras or an elevated balustrade that could hypothetically allow people to observe voters in the voting booths.

- Electoral System: includes complaints relating to specific aspects of the German electoral system (overhang mandates, the 5% threshold, the method used for turning votes into seats, etc.) Also includes complaints that do not criticize a specific aspect of the German electoral system, rather a broader issue that is related to the electoral system.

- Party List Not on Ballot/Other Ballot Access Issues: many parties whose party lists were not recognized by the Bundeswahlleiter (and therefore did not appear as options under the second vote) complained about this impediment. This category also includes complaints from independent candidates about their placement on the ballot.

- Problems with the creation of Party Lists: cases where complainants claim that party lists were improperly prepared. In Germany, the candidates and their order on the individual Landlists are determined by the parties themselves at a mass gathering of each party.

- Counting of the votes: any complaint that alleges inconsistencies in vote counts or improper procedures in the preparation of those counts.

- Improper Statistics/Representative Election Statistic: any complaint that alleges the violation of the secret ballot through the preparation of certain election statistics. In Germany, the most salient of these statistics is the “Repräsentative Wahlstatistik” (Representative Election Statistic). This determines the voting patterns of Germans differentiated by sex and age range, which is accomplished through sampling precincts throughout Germany by having them distribute marked ballots (these ballots indicate the voter’s sex and age-range). This process is controversial, as numerous complainants objected to the perceived invasion of privacy.

- Improper District Boundaries: indicates a complaint that alleged improprieties in the drawing of district boundaries.

- Allegations of official corruption: complaints that accuse various government officials of involvement in various corruption schemes (this does not include allegations of improprieties against poll workers).

- Unspecified Other: includes complaints where the nature of the complaint could not be ascertained or non-sequitur complaints.

Many of these codes are closely related (for example, many complaints coded under “Absentee-ballot related problem” involve problems with voter registration and as such could also be coded as “Registration related problems”). The same issue presents itself with many of the codes developed specifically for the German case, as they deal with numerous specific complaints about the electoral system.
To determine how to code the “reason” for the complaint, we consulted the “Betreff” (subject) field that is contained in the original Bundestag files, in the table of contents alongside the corresponding file number. This field gives an approximation of the nature of the complaint as parsed by the committee. The documents also contain the specifics of every complaint as well as the response of the committee. As such, both the complaints and the committee’s responses can be quite lengthy. The four documents—17/2250, 17/3100, 17/4600 and 17/6300—that relate to the 2009 Bundestagswahl are 56, 212, 136, and 144 pages long respectively. The reason codes in the database take into account both the subject of the complaint as assigned in the table of contents and the broader enumeration of the complaint found in the body of the document.

Some complaints are assigned multiple “reason” codes (up to six). This is usually precipitated by a telltale “u.a.” (“unter anderem,” meaning among others) in the original “Betreff” of the complaint. The precise nature of each multifaceted complaint is completely enumerated in the main text of each complaint, as opposed to the “Betreff.” In order to receive multiple codes, a complaint had to enumerate multiple complaints that involved multiple EIRS+ codes, not simply multiple aspects of the same code. For example, in 2009 many people complained about overhang mandates, a complaint that was coded under “Electoral System.” Many people also complained about the distribution of seats, which would also fall under an “Electoral System” complaint. In some cases, these complaints involved both the division of seats and overhang mandates, in which case the complaint was still simply coded as a single “Electoral System” complaint, regardless of the fact that there are two complaints about the electoral system. The codes are designed to show the subjects of the complaints, as opposed to their multiplicity.

We combine several of the distinct complaint codes into composite categories. In some cases, certain codes can be considered subsets of other codes. In other cases, the codes are intrinsically related enough to justify their combination. We combine the codes “Polling Place Problem,” “Disability Access Problem,” “Absentee Ballot Related Problem” and “Registration Related Problem” because they all relate generally to a voter encountering difficulty casting, or being outright denied, their vote, whether through problems in the administration of the Briefwahl to in-person problems at the Wahllokal (precinct). We combine the “Electoral System” and “Problem with the creation of Party Lists” codes due to the similar phenomena they describe (the “Problem with the creation of Party Lists” is a specific aspect of the electoral system that plays a particularly important role in Germany). The desire to create a unified structure for the subjects of these complaints within the model motivated these combinations. By combining complaints that deal with similar problems in election administration, the independence of the distinctly coded disputes increases.

To determine the district from which the complaint emanated, we used the postal codes given in the documents from the Bundestag in conjunction with several sets of shapefiles (one that shows the distribution of postal codes across Germany and the other that shows the division of the electoral districts in a given Bundestag election). It should be noted that the locations given in the files are only the location of the complainant, i.e., it is entirely possible for someone to complain about an issue that they themselves did not experience—standing is not an issue.

In the analysis that relates the occurrence of complaints to the Klimek et al. (2012)
model fraud probability parameters, we treat the complaints and their location in a binary way. The types of disputes are aggregated to the district level: if there is a positive number of complaints of a certain type then the whole district receives a “1” for that type, and if there are no complaints of a certain type in a district then it receives a “0”. Despite the fact that there are districts with more than one complaint of the same type, we decided that considering the multiplicity of a particular type of complaint in a given district would be ill-advised for several reasons. The first is that the complaint data was coded not to represent the number of complaints of a certain type in a district, but to illuminate the type of complaints levied by citizens in a given district. In many cases, determining the precise number of disputes of the same type contained in a given complaint is a nebulous task, as often these disputes are inextricably related to one another. This is especially true of complaints that allege the unconstitutionality of various aspects of the electoral system. Second, considering the multiplicity of complaints could bias the analysis in favor of well-publicized problems with electoral administration. Just because more people filed complaints of a given nature does not mean that these complaints are more valid or serious than a complaint filed by just one individual. One example where the preceding logic would be less applicable would be in a situation where many individuals allege that their votes had not been counted or that they had otherwise been illegally barred from voting. This is not the case regarding the German disputes, however. Third, even in districts where relatively more complaints originated, the number of complainants still represents a tiny percentage of the population and reading too much into the sheer number of complaints risks distorting the resulting model. Despite the possible benefits of looking at the multiplicity of complaints given certain requirements, the 2009 complaint data from Germany do not meet these requirements.

Mexican nullification petitions data:

In Mexico, the General Law Of Methods Regarding Disputed Elections, passed on 22 November 1996 as part of a series of electoral reforms, allows representatives of the various political parties to petition for the results of a disputed casilla to be nullified (Cámara de Diputados del H. Congreso de la Unión 2014). For our purposes, there are three relevant articles in this Law.

Article 9 of the General Law describes a broad set of of criteria that must be included in any petition submitted to the Election Tribunal of the Federal Judicial Branch (TEPJF), such as an enumeration of the complaint in an “expressed and clear” manner and a valid signature from an appropriate political party representative.

Article 52 of the same document deals with a more specific set of criteria that must be included in petitions to have the results of various casillas nullified. Among the five criteria that must be included in each petition are the following:

- The election being disputed (President, Senate, etc.)
- An indication of the entity whose election result certification is being challenged
- An indication of precisely which casillas are being challenged and the cause for the challenge (see below)
• An indication of any arithmetic errors present in the count being challenged, if appropriate

• The connectedness of this challenge to others in the same election, if appropriate

Additionally, Article 75, paragraph 1 of the General Law lists eleven official irregularities that, if found to be credible, are grounds for a casilla to be declared null, and only one must be found credible. They include:

• A) Installing a casilla in a location other than that indicated for use by the corresponding district council, without just cause;

• B) Submitting packets containing District voting records outside the time that the Federal Code of Institutions and Electoral Procedures permits, without just cause;

• C) Carrying out computation and scrutiny of votes in a location other than the one approved by the corresponding district council, without just cause;

• D) Receiving votes at a date or time other than during approved hours on Election Day;

• E) Having votes collected by people other than those charged by the Federal Code of Electoral Institutions and Procedures with collecting votes;

• F) Willful misconduct or error in the vote count, given that the act is determinative for the outcome of the vote;

• G) Allowing citizens to vote without presenting proper credentials or without being present on the list of registered voters (except in cases permitted under the Federal Code of Institutions and Electoral Procedures as well as article 85 of this document), given that the act is determinative for the outcome of the vote;

• H) Impeding the ability of political party representatives to observe elections at the polling place, or expelling them, without just cause;

• I) Pressure or physical violence exacted on voters or polling place workers, given that the act is determinative for the outcome of the vote;

• J) Impeding the ability of citizens to exercise their right to vote, without just cause and given that the act is determinative for the outcome of the vote;

• K) Existence of serious, fully accredited irregularities that were not repairable during the election cycle nor during the computation and scrutiny of votes, that put in doubt the result of the vote and that are determinative for the outcome of the vote.

It should be noted that of the 375 documents published regarding nullification petitions during the 2006 presidential election, approximately ten percent did not contain usable data because the petitioner failed to comply with the requirements stated in one or more of Articles 9, 52 and 75. In these cases, the petitioner typically failed either to enumerate
precisely which casillas he or she intended to challenge or to list an Article 75 cause for the nullification request, leaving the TEPJF judges with no choice but to discard the incomplete petition. The document SUP-JIN-00004-2006.htm\(^{43}\) is an example of this.

**Data Collection Methods:** We examined the 375 documents published by the TEPJF on 28 August 2006 that contain the Tribunal’s decisions regarding all nullification petitions connected with the 2 July 2006 Presidential Election.\(^{44}\) We extracted data on 27,109 challenged casillas, including the district and sección numbers, casilla type, and the grounds for each petition under Article 75 of the General Law. These data primarily came from summary tables provided in the petitions that show casilla number and type with an X in each cell corresponding to the Article 75 provision (A–K) under which the casilla was challenged. Many casillas were challenged using multiple provisions. Other data, when not listed in tables, were extracted from the text of the documents by reading through each one.


\(^{44}\)The decisions are posted in files at URLs like http://portal.te.gob.mx/colecciones/sentencias/html/SUP/2006/JIN/SUP-JIN-00XXX-2006.htm where XXX ∈ \{001,...,375\}.
References


Ochoa, Alejandra Loera. 2000. “Estudio comparado de las nulidades del sistema de medios


Table 1: Fraud Probability Parameter Estimates

<table>
<thead>
<tr>
<th>Type</th>
<th>Original</th>
<th>Chi-squared</th>
<th>Mixture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( f_i )</td>
<td>( f_e )</td>
<td>( \alpha )</td>
</tr>
<tr>
<td>Russia, 2011, regions</td>
<td>.158</td>
<td>.029</td>
<td>4.96</td>
</tr>
<tr>
<td>Russia, 2011, territories</td>
<td>.980</td>
<td>.003</td>
<td>3.19</td>
</tr>
<tr>
<td>Russia, 2011, precincts(^a)</td>
<td>.998</td>
<td>.004</td>
<td>3.42</td>
</tr>
<tr>
<td>Mexico, 2012, districts</td>
<td>.008</td>
<td>0</td>
<td>5.34</td>
</tr>
<tr>
<td>Michigan, 2010, counties</td>
<td>.200</td>
<td>.015</td>
<td>5.86</td>
</tr>
<tr>
<td>California, 2008, precincts(^a)</td>
<td>.996</td>
<td>0</td>
<td>1.71</td>
</tr>
</tbody>
</table>

Note: \( f_i, f_e \) and \( \alpha \) estimates using the original, chi-squared and finite mixture versions of the Klimek et al. (2012) model. \( n = 84 \) Russian regions; \( n = 2718 \) Russian territories; \( n = 500 \) or \( n = 1000 \) Russian or California precincts; \( n = 300 \) Mexican districts; \( n = 83 \) Michigan counties. \(^a\) original and chi-squared estimates use samples of 1,000 precincts, and the mixture estimates use samples of 500 precincts. \(^b\) parameter not identified when \( f_i = f_e = 0. \)
Table 2: Estimates and Bootstrap Standard Errors for the Finite Mixture Model

<table>
<thead>
<tr>
<th>Election</th>
<th>Parameter</th>
<th>MLE</th>
<th>SE</th>
<th>ubound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russia, 2011 regions</td>
<td>( f_i )</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>( f_e )</td>
<td>.1359</td>
<td>.0385</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>( \alpha )</td>
<td>14.025</td>
<td>4.187</td>
<td>( -^b )</td>
</tr>
<tr>
<td></td>
<td>( a )</td>
<td>.5564</td>
<td>.0055</td>
<td>.5566</td>
</tr>
<tr>
<td></td>
<td>( \nu )</td>
<td>.2950</td>
<td>.0060</td>
<td>.4453</td>
</tr>
<tr>
<td></td>
<td>( \sigma_\nu )</td>
<td>.0735</td>
<td>.0134</td>
<td>.0735</td>
</tr>
<tr>
<td></td>
<td>( \sigma_a )</td>
<td>.0510</td>
<td>.0043</td>
<td>.0510</td>
</tr>
<tr>
<td></td>
<td>( \theta )</td>
<td>( -^a )</td>
<td>( -^a )</td>
<td>.7979</td>
</tr>
</tbody>
</table>

Note: \( n = 84 \); loglikelihood, \(-2196.7\). MLE, maximum likelihood estimate; SE, parametric bootstrap standard error (150 bootstrap samples); ubound, parameter upper bound.

\( ^a \) parameter not identified when \( f_i = 0 \). \( ^b \) unbounded.
Table 3: Frequency of Postelection Complaint Types, Germany 2009

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allegations</td>
<td>Allegations of official corruption</td>
<td>3</td>
<td>1.83</td>
</tr>
<tr>
<td>Ballot</td>
<td>Ballot related problem</td>
<td>3</td>
<td>1.83</td>
</tr>
<tr>
<td>Campaigning</td>
<td>Improper Campaigning Influence</td>
<td>15</td>
<td>9.15</td>
</tr>
<tr>
<td>Counting</td>
<td>Counting of the votes</td>
<td>6</td>
<td>3.66</td>
</tr>
<tr>
<td>Criminal</td>
<td>Criminal status related problem</td>
<td>3</td>
<td>1.83</td>
</tr>
<tr>
<td>District</td>
<td>Improper District Boundaries</td>
<td>2</td>
<td>1.22</td>
</tr>
<tr>
<td>ID</td>
<td>Identification related problem</td>
<td>13</td>
<td>7.93</td>
</tr>
<tr>
<td>Other</td>
<td>Unspecified Other</td>
<td>5</td>
<td>3.05</td>
</tr>
<tr>
<td>Party List</td>
<td>Party List Not on Ballot</td>
<td>22</td>
<td>13.41</td>
</tr>
<tr>
<td>Polling Place</td>
<td>Absentee, Registration, Disability, Polling place</td>
<td>28</td>
<td>17.07</td>
</tr>
<tr>
<td>Statistics</td>
<td>Improper Statistics</td>
<td>8</td>
<td>4.88</td>
</tr>
<tr>
<td>System</td>
<td>Electoral System, Party Lists creation</td>
<td>56</td>
<td>34.15</td>
</tr>
</tbody>
</table>

Note: Number of districts that have each type of complaint. Source: Compiled from archives of the Bundestag’s website for the Ausschuss für Wahlprüfung, Immunität und Geschäftsordnung (see the Appendix). Types are described in the Appendix.

Table 4: Residual Deviance Statistics for Various Complaints Models, Germany 2009

<table>
<thead>
<tr>
<th>Type</th>
<th>null (df)</th>
<th>chi-squared (df)</th>
<th>original (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polling Place</td>
<td>185.91 (298)</td>
<td>175.26 (294)</td>
<td>182.97 (297)</td>
</tr>
<tr>
<td>Counting</td>
<td>58.783 (298)</td>
<td>54.043 (297)</td>
<td>—a</td>
</tr>
<tr>
<td>Criminal</td>
<td>33.581 (298)</td>
<td>28.923 (294)</td>
<td>—a</td>
</tr>
<tr>
<td>Campaigning</td>
<td>119.01 (298)</td>
<td>114.27 (294)</td>
<td>—a</td>
</tr>
<tr>
<td>Statistics</td>
<td>73.720 (298)</td>
<td>66.121 (294)</td>
<td>70.799 (297)</td>
</tr>
<tr>
<td>Party List</td>
<td>157.15 (298)</td>
<td>153.64 (297)</td>
<td>—a</td>
</tr>
</tbody>
</table>

Note: null model df, 298. a there are no significant terms involving $f_i$, $f_o$ or $\alpha$. 
Table 5: Effects of Election Fraud Probabilities on Postelection Complaints, Germany 2009

<table>
<thead>
<tr>
<th>Type</th>
<th>Regressor</th>
<th>coef.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polling Place</td>
<td>Intercept</td>
<td>-7.257e-01</td>
<td>9.958e-01</td>
</tr>
<tr>
<td></td>
<td>$f_i$</td>
<td>-4.443e+02</td>
<td>1.874e+02</td>
</tr>
<tr>
<td></td>
<td>$f_e$</td>
<td>-5.868e+02</td>
<td>3.322e+02</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>-1.384e-01</td>
<td>2.867e-01</td>
</tr>
<tr>
<td></td>
<td>$f_i f_e$</td>
<td>1.873e+05</td>
<td>6.768e+04</td>
</tr>
<tr>
<td>Counting</td>
<td>Intercept</td>
<td>-2.246</td>
<td>0.828</td>
</tr>
<tr>
<td></td>
<td>$f_i$</td>
<td>-614.663</td>
<td>351.514</td>
</tr>
<tr>
<td>Criminal</td>
<td>Intercept</td>
<td>-0.3198</td>
<td>3.0159</td>
</tr>
<tr>
<td></td>
<td>$f_i$</td>
<td>-657.4213</td>
<td>653.2632</td>
</tr>
<tr>
<td></td>
<td>$f_e$</td>
<td>506.9327</td>
<td>332.2950</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>-2.8229</td>
<td>1.7043</td>
</tr>
<tr>
<td></td>
<td>$f_i \alpha$</td>
<td>321.1325</td>
<td>272.1003</td>
</tr>
<tr>
<td>Campaining</td>
<td>Intercept</td>
<td>-4.636</td>
<td>1.237</td>
</tr>
<tr>
<td></td>
<td>$f_i$</td>
<td>4.331e+02</td>
<td>2.147e+02</td>
</tr>
<tr>
<td></td>
<td>$f_e$</td>
<td>8.319e+02</td>
<td>3.883e+02</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>4.688e-02</td>
<td>3.925e-01</td>
</tr>
<tr>
<td></td>
<td>$f_i \alpha$</td>
<td>-2.329e+05</td>
<td>1.245e+05</td>
</tr>
<tr>
<td>Statistics</td>
<td>Intercept</td>
<td>1.5621</td>
<td>1.9034</td>
</tr>
<tr>
<td></td>
<td>$f_i$</td>
<td>-251.9137</td>
<td>225.1443</td>
</tr>
<tr>
<td></td>
<td>$f_e$</td>
<td>-3068.2959</td>
<td>1437.0653</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>-1.5874</td>
<td>0.7778</td>
</tr>
<tr>
<td></td>
<td>$f_i \alpha$</td>
<td>1076.9818</td>
<td>488.3215</td>
</tr>
<tr>
<td>Party List</td>
<td>Intercept</td>
<td>-1.8486</td>
<td>0.4236</td>
</tr>
<tr>
<td></td>
<td>$f_i$</td>
<td>-209.3588</td>
<td>124.0320</td>
</tr>
</tbody>
</table>

Note: binomial regression (logit) model coefficient estimates; $n = 299$. $f_i$, $f_e$ and $\alpha$ estimates are from the chi-squared Klimek et al. (2012) model.
Table 6: Frequency of Nullification Petition Claim Types, Mexico 2006

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Installing a ballot box in the wrong location</td>
<td>639</td>
<td>1.51</td>
</tr>
<tr>
<td>B</td>
<td>Submitting district voting records late</td>
<td>1,612</td>
<td>3.82</td>
</tr>
<tr>
<td>C</td>
<td>Counting the votes in the wrong location</td>
<td>397</td>
<td>.94</td>
</tr>
<tr>
<td>D</td>
<td>Receiving votes during unapproved times</td>
<td>1,307</td>
<td>3.10</td>
</tr>
<tr>
<td>E</td>
<td>Having unapproved poll workers collect votes</td>
<td>6,850</td>
<td>16.23</td>
</tr>
<tr>
<td>F</td>
<td>Willful misconduct or error in the vote count</td>
<td>21,116</td>
<td>50.02</td>
</tr>
<tr>
<td>G</td>
<td>Allowing citizens to vote without proper credentials</td>
<td>2,061</td>
<td>4.88</td>
</tr>
<tr>
<td>H</td>
<td>Impeding observation by party representatives</td>
<td>150</td>
<td>.36</td>
</tr>
<tr>
<td>I</td>
<td>Pressure or physical violence at the polling place</td>
<td>1,871</td>
<td>4.43</td>
</tr>
<tr>
<td>J</td>
<td>Impeding the ability of citizens to vote</td>
<td>660</td>
<td>1.56</td>
</tr>
<tr>
<td>K</td>
<td>Other serious irregularities</td>
<td>5,555</td>
<td>13.16</td>
</tr>
</tbody>
</table>

Note: Number of casillas that have each type of petition claim type. Source: Compiled from documents posted in files at URLs like http://portal.te.gob.mx/colecciones/sentencias/html/SUP/2006/JIN/SUP-JIN-00XXX-2006.htm where XXX ∈ {001,...,375}. Types are described in the Appendix.

Table 7: Correlations Among Different Models’ Fraud Parameter Estimates, Mexico 2006

<table>
<thead>
<tr>
<th>Model and data description</th>
<th>$\chi^2$</th>
<th>M-s</th>
<th>M-c</th>
<th>M-P0</th>
<th>M-P1</th>
</tr>
</thead>
<tbody>
<tr>
<td>chi-squared, sección, deputies:</td>
<td>$\chi^2$</td>
<td>0.20</td>
<td>0.20</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>original, sección, deputies:</td>
<td>M-s</td>
<td>−0.01</td>
<td>0.38</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td>original, casilla, deputies:</td>
<td>M-c</td>
<td>0.10</td>
<td>0.31</td>
<td>0.07</td>
<td>0.14</td>
</tr>
<tr>
<td>original, sección, president, PAN winner:</td>
<td>M-P0</td>
<td>0.51</td>
<td>−0.03</td>
<td>0.08</td>
<td>0.55</td>
</tr>
<tr>
<td>original, sección, president, district winner:</td>
<td>M-P1</td>
<td>0.48</td>
<td>0.06</td>
<td>0.10</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Note: product-moment correlations; upper triangle has correlations for $f_i$; lower triangle has correlations for $f_e$. Labels in the left column describe the model, data aggregation level and office.
Table 8: Residual Deviance Statistics for Various Nullification Petition Models, Mexico 2006

<table>
<thead>
<tr>
<th>Type</th>
<th>null chi-squared</th>
<th>Deputy Votes</th>
<th>Presidential Votes</th>
<th>original&lt;sup&gt;a&lt;/sup&gt;</th>
<th>original&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6966.1</td>
<td>6685.6</td>
<td>6876.1</td>
<td>6773.4</td>
<td>6907.4</td>
</tr>
<tr>
<td>B</td>
<td>14682</td>
<td>13578</td>
<td>13729</td>
<td>12450</td>
<td>13519</td>
</tr>
<tr>
<td>C</td>
<td>4605.0</td>
<td>4194.8</td>
<td>4224.2</td>
<td>4282.9</td>
<td>4210.0</td>
</tr>
<tr>
<td>D</td>
<td>12420</td>
<td>11902</td>
<td>12107</td>
<td>11700</td>
<td>11947</td>
</tr>
<tr>
<td>E</td>
<td>42599</td>
<td>41860</td>
<td>41421</td>
<td>41903</td>
<td>42008</td>
</tr>
<tr>
<td>F</td>
<td>91174</td>
<td>90202</td>
<td>90474</td>
<td>88974</td>
<td>90345</td>
</tr>
<tr>
<td>G</td>
<td>18719</td>
<td>17563</td>
<td>17850</td>
<td>17420</td>
<td>17614</td>
</tr>
<tr>
<td>H</td>
<td>1902.7</td>
<td>1645.2</td>
<td>1865.2</td>
<td>1865.8</td>
<td>1782.3</td>
</tr>
<tr>
<td>I</td>
<td>18488</td>
<td>17341</td>
<td>15986</td>
<td>14177</td>
<td>14822</td>
</tr>
<tr>
<td>J</td>
<td>6995.2</td>
<td>6597.1</td>
<td>6026.5</td>
<td>6712.9</td>
<td>6539.4</td>
</tr>
<tr>
<td>K</td>
<td>40773</td>
<td>37944</td>
<td>39121</td>
<td>39494</td>
<td>39087</td>
</tr>
</tbody>
</table>

Note: \( n = 64,689 \); null model df, 64,685; other models df, 64,678. \(^a\) using PAN as the district winner except in five districts. \(^b\) using the party with the most votes in each district as the district’s “winner”
Figure 1: Proportions of Votes Shifted by Frauds, Russia 2011

Region Data

Incremental Fraud
Nonvoters

Extrem Fraud
Nonvoters

Opposition
Opposition

Note: along the $x$-axis are the proportions of votes shifted to winner given either incremental or extreme fraud based on finite mixture model estimates; the $y$-axis shows the probability density for the proportion values.
Figure 2: Incremental Fraud Probabilities, by District, Germany 2009 Bundestag Erststimmen

Note: $f_i$ values estimated using the chi-squared variant of the Klimek et al. (2012) model, separately for each district using sección data. Color red means $f_i / \max(f_i) = 1$, color blue means $f_i = 0$, and intermediate values of $f_i$ have colors that are weighted mixtures of red and blue. $\max(f_i) = 0.0162$. 
Figure 3: Extreme Fraud Probabilities, by District, Germany 2009 Bundestag Erststimmen

Note: $f_e$ values estimated using the chi-squared variant of the Klimek et al. (2012) model, separately for each district using sección data. Color red means $f_e / \max(f_e) = 1$, color blue means $f_e = 0$, and intermediate values of $f_e$ have colors that are weighted mixtures of red and blue. \( \max(f_e) = 0.0087 \).
Figure 4: Postelection Complaint Type Probabilities, Germany 2009

(a) Polling place, large alpha

(b) Polling place, small alpha

(c) Campaigning, large alpha

(d) Campaigning, small alpha

(e) Statistics, large alpha

(f) Statistics, small alpha

Note: nonparametric regression of probabilities simulated from binomial regression model estimates using the chi-squared Klimek et al. (2012) model estimates for $f_i$, $f_e$ and $\alpha$; “large alpha” means $\alpha > \text{median}(\alpha)$, and “small alpha” means $\alpha < \text{median}(\alpha)$. 
Figure 5: Postelection Complaint Type Probabilities, Germany 2009

Note: Solid line shows predicted values of probabilities simulated from binomial regression model estimates using the chi-squared Klinek et al. (2012) model estimates for \( f_i \); dashed lines show 95% confidence bounds, and rug plot shows the locations of the \( f_i \) estimates.
Note: cross-validation comparison of binomial regression models using the original and the chi-squared Klimek et al. (2012) model estimates for $f_i$, $f_e$ and $\alpha$, as well as random parameter values.
Figure 7: Indicators for Voters' Strategic Behavior, Germany 2009, CDU-CSU

Note: nonparametric regression contours for \( f_e \). “[party]: (Zweit.-Erst.)/ballots” is the total of \( \text{Zweitstimmen} \) cast for [party] minus the number of \( \text{Erststimmen} \) cast for [party] divided by the total number of ballots used in the district. “margin: 1 versus 3” is the number of \( \text{Erststimmen} \) for the winning party in each district minus the number of votes for the third-place party divided by the total of \( \text{Erststimmen} \) cast in the district. Rug plots show locations of district values. \( p \) in each subfigure heading reports the \( p \)-value for a significance test versus the model of no effects.
Figure 8: Indicators for Voters’ Strategic Behavior, Germany 2009, FDP

Note: nonparametric regression contours for $f_e$. “[party]: (Zweit.-Erst.)/ballots” is the total of Zweitstimmen cast for [party] minus the number of Erststimmen cast for [party] divided by the total number of ballots used in the district. “margin: 1 versus 3” is the number of Erststimmen for the winning party in each district minus the number of votes for the third-place party divided by the total of Erststimmen cast in the district. Rug plots show locations of district values. $p$ in each subfigure heading reports the $p$-value for a significance test versus the model of no effects.
Figure 9: Indicators for Voters’ Strategic Behavior, Germany 2009, SPD

Note: nonparametric regression contours for \( f_e \). “[party]: (Zweit.-Erst.)/ballots” is the total of Zweitstimmen cast for [party] minus the number of Erststimmen cast for [party] divided by the total number of ballots used in the district. “margin: 1 versus 3” is the number of Erststimmen for the winning party in each district minus the number of votes for the third-place party divided by the total of Erststimmen cast in the district. Rug plots show locations of district values. \( p \) in each subfigure heading reports the \( p \)-value for a significance test versus the model of no effects.
Note: nonparametric regression contours for $f_e$. “[party]: (Zweit.-Erst.)/ballots” is the total of Zweitstimmen cast for [party] minus the number of Erststimmen cast for [party] divided by the total number of ballots used in the district. “margin: 1 versus 3” is the number of Erststimmen for the winning party in each district minus the number of votes for the third-place party divided by the total of Erststimmen cast in the district. Rug plots show locations of district values. $p$ in each subfigure heading reports the $p$-value for a significance test versus the model of no effects.
Figure 11: Indicators for Voters’ Strategic Behavior, Germany 2009, Die Linke

(a) $fe: p = 0.0075$

(b) $fe: p = 0.01$

(c) $fe: p = 1e^{-04}$

(d) $fe: p = 0.0083$

Note: nonparametric regression contours for $f_e$. “[party]: (Zweit.-Erst.)/ballots” is the total of Zweitstimmen cast for [party] minus the number of Erststimmen cast for [party] divided by the total number of ballots used in the district. “margin: 1 versus 3” is the number of Erststimmen for the winning party in each district minus the number of votes for the third-place party divided by the total of Erststimmen cast in the district. Rug plots show locations of district values. $p$ in each subfigure heading reports the $p$-value for a significance test versus the model of no effects.
Note: $f_i$ values estimated using the chi-squared variant of the Klimek et al. (2012) model, separately for each district using sección data. Color red means $f_i = 1$, color blue means $f_i = 0$, and intermediate values of $f_i$ have colors that are weighted mixtures of red and blue.
Figure 13: Extreme Fraud Probabilities, by District, Mexico 2006 House of Delegates

Note: \( f_e \) values estimated using the chi-squared variant of the Klimek et al. (2012) model, separately for each district using sección data. Color red means \( f_e / \max(f_e) = 1 \), color blue means \( f_e = 0 \), and intermediate values of \( f_e \) have colors that are weighted mixtures of red and blue. \( \max(f_e) = 0.0105 \).
Figure 14: Incremental Fraud Probabilities, by District, Mexico 2006 President

Note: $f_i$ values estimated using the original Klimek et al. (2012) model, separately for each district using sección data. Color red means $f_i = 1$, color blue means $f_i = 0$, and intermediate values of $f_i$ have colors that are weighted mixtures of red and blue.
Figure 15: Extreme Fraud Probabilities, by District, Mexico 2006 President

Note: $f_e$ values estimated using the original Klimek et al. (2012) model, separately for each district using sección data. Color red means $f_e / \max(f_e) = 1$, color blue means $f_e = 0$, and intermediate values of $f_e$ have colors that are weighted mixtures of red and blue. $\max(f_e) = 0.024$. 
Figure 16: Nullification Petition Complaint Type Probabilities, Mexico 2006

Note: nonparametric regression of simulated probabilities; “large alpha” means \( \alpha > \text{median}(\alpha) \), and “small alpha” means \( \alpha < \text{median}(\alpha) \).
Figure 17: Nullification Petition Complaint Type Probabilities, Mexico 2006

Note: nonparametric regression of simulated probabilities; “large alpha” means $\alpha > \text{median}(\alpha)$, and “small alpha” means $\alpha < \text{median}(\alpha)$. 
Figure 18: Predicting Nullification Petitions, Mexico 2006

Note: cross-validation comparison of overdispersed binomial regression models using the original and the chi-squared Klimek et al. (2012) model estimates for $f_i$, $f_e$ and $\alpha$, the models based on presidential votes where PAN is always the winner and where the winner varies by district, and the model that includes only the covariates.