Detecting Attempted Election Theft: Vote Counts, Voting Machines and Benford's Law *

Walter R. Mebane, Jr.†

April 19, 2006

*Prepared for presentation at the 2006 Annual Meeting of the Midwest Political Science Associate, Chicago, IL, April 20–23. Thanks to Daniel Dauplaise for sparking my interest in Benford's Law, and to Charlie Gibbons for several helpful suggestions.

†Professor, Department of Government, Cornell University, 217 White Hall, Ithaca, NY 14853–7901 (Phone: 607-255-2868; Fax: 607-255-4530; E-mail: wrm1@cornell.edu).

Fraudulent elections and disputes about election outcomes are nothing new. Gumbel (2005) reviews the sorry history of deceit and electoral manipulation in America, going back to the dawn of the republic. Throughout the world, in old and new democracies alike, allegations of vote fraud frequently occur (Lehoucq 2003). One new element is voting technologies that make some familiar methods for physically verifying the accuracy of vote totals impossible to use. The advent of electronic voting machines means that often now there are no paper ballots to be recounted. To steal an election it is no longer necessary to toss boxes of ballots in the river, stuff the boxes with thousands of phony ballots, or hire vagrants to cast repeated illicit votes. All that may be needed nowadays is access to an input port and a few lines of computer code. To detect such manipulations is a difficult and urgent problem. In terms of legitimacy it is not clear whether the worse problem is that erroneous election outcomes may occur or that many may not believe that correct outcomes are valid.

This paper introduces statistical methods intended to help detect election fraud. Other methods, using regression-based techniques for outlier detection, have previously been proposed to help detect election anomalies (e.g. Wand, Shotts, Sekhon, Mebane, Herron, and Brady 2001; Mebane, Sekhon, and Wand 2001). The methods described here are distinctive in that they do not require that we have covariates to which we may reasonably assume the votes are related across political jurisdictions. For one set of methods I describe—methods based on tests of the distribution of the digits in reported vote counts—all that is needed are the vote counts themselves. I study the application of those methods to both precinct-level and voting machine-level vote tabulations. Part of the potential practical relevance of these methods is that situations in which little more than the vote counts are available may arise frequently in connection with actual election controversies.

The other set of methods I describe, which are based on testing whether votes are randomly assigned to the voting machines used for a voting precinct, require candidate vote totals disaggregated to the level of individual voting machines. More than that, these methods also require that a fair amount is known about how the voting machines were used. For instance, for voting machines used during early voting periods,[1] we need to know on which days particular

[1]See Gronke, Bishin, Stevens, and Galanes-Rosenbaum (2005) for a discussion of early voting in Florida during the 2004 election.

machines were used, and at which early voting site. In fact it may be useful to know the exact time at which each vote was cast and on which machine. Such details are routinely available when some kinds of electronic voting machines are used, except that it may not be possible to tell when a particular vote was cast: transaction event logs maintained for each machine indicate when a vote was cast, but to help protect the secret ballot it is not possible to match an individual vote record (an individual ballot image) with a particular transaction.

Both methods depend in different ways on ideas about voter behavior. The methods that check whether votes are randomly assigned to machines assume that voters' choices between candidates do not depend on the particular voting machine they use. If a set of machines are all used in the same precinct during the same period of time, and yet the distribution of vote choices varies significantly across machines, then the idea is to attribute the variation to some kind of manipulation. Perhaps voters with different preferences were somehow directed to use different machines. Or perhaps some of the machines were hacked.

The methods that check the distribution of the digits in reported vote counts depend on ideas about voter choice behavior that differ substantially from the models usually used in research on political behavior. The digit-test methods are based on the expectation that the second digits of vote counts should satisfy Benford's Law (Hill 1995). Benford's Law specifies that the ten possible second digits should not occur with equal frequency. A fundamental question is why we should expect Benford's Law to apply to vote count data. Even though some have proposed to use the second-digit Benford's Law distribution to test for fraudulent votes (e.g., Pericchi and Torres 2004), prominent election monitors have strongly disputed such proposals (Carter Center 2005). I suggest that a behavioral focus on the individualized uncertainty in each person's vote choice may be inappropriate when thinking about vote counts for the purpose of trying to decide whether the counts are fraudulent. Indeed, leaving aside questions of vote fraud, to the extent that the familiar kinds of behavioral models cannot in general produce vote counts with second digits that follow the Benford's Law distribution—and, in general, they cannot—the fact that vote counts do often satisfy Benford's Law is strong evidence that the familiar behavioral models do not describe the votes people actually cast.

Even if Benford's Law typically describes vote count data, it does not follow that deviations from Benford's Law indicate election fraud. I present the results of some simulation exercises that begin exploring what if any kinds of vote fraud a test based on the second-digit Benford's Law distribution can detect. In the limited range of simulations I have conducted so far, I find that the Benford's Law test is sensitive to some kinds of manipulation of vote counts but not to others. The test seems sensitive enough to warrant further exploration of its properties. I think it has an excellent chance of developing into a standard tool for forensically auditing elections.

I apply both the vote randomization test and the Benford's Law test to data from three Florida counties in the 2004 general election. The available data include ballot image and voting machine event log files for electronic early voting and electronic polling place votes in Broward, Miami-Dade and Pasco counties, including labels identifying the precinct and voting machine for each ballot.[2]

A Randomization Test for Voting Machines

The first test addresses whether the distribution of the votes is the same on all of each precinct's voting machines. The idea is to assess whether the votes cast in each precinct were randomly and independently assigned to each machine used in the precinct. A manipulation of the vote that affected some machines but not others would probably cause the distribution of the votes among candidates to differ on the affected machines. Testing that the split of the votes is the same on all the machines used in a precinct is one way to check for such selective manipulation. Voter preferences vary substantially from precinct to precinct, but if a collection of machines is used to count the votes in a precinct, with all of the machines being used throughout the same period of time, and if each voter has the same probability of being assigned to each machine, then the split of the votes should be roughly the same on all of the precinct's machines.

To define the test, let π_{ijk} denote the probability that voter i in precinct j is assigned to vote using machine k, and let ρ_{ijkl} denote the probability that voter i in precinct j using machine k chooses candidate l. The number of voters in precinct j is n_j, the number of machines is m_j, and

[2]The ballot image and event log file data were supplied by David Dill. Additional data regarding characteristics of the machines used in Miami-Dade were supplied by Martha Mahoney. For more information about data sources see the Data Note.

$\sum_{k=1}^{m_j} \pi_{ijk} = 1$. The number of votes expected for candidate l in precinct j on machine k is

$$V_{jkl} = \sum_{i=1}^{n_j} \pi_{ijk}\rho_{ijkl} ,$$

and the expected vote share for candidate l in precinct j on machine k is

$$R_{jkl} = \left(\sum_{i=1}^{n_j} \pi_{ijk}\right)^{-1} V_{jkl} .$$

If the probability of being assigned to a machine is the same for each voter in precinct j, then $\pi_{ijk} = \pi_{jk}$. If neither the choice the voter makes nor the choice that is recorded depends on either the machine or on how other voters are assigned to machines, then $\rho_{ijkl} = \rho_{ijl}$. If both of these conditions hold, the number of votes expected for candidate l in precinct j on machine k is

$$\tilde{V}_{jkl} = \pi_{jk} \sum_{i=1}^{n_j} \rho_{ijl} ,$$

and the expected machine vote share is

$$\tilde{R}_{jkl} = \frac{\tilde{V}_{jkl}}{n_j \pi_{jk}} = n_j^{-1} \sum_{i=1}^{n_j} \rho_{ijl} .$$

In this case the vote share expected for candidate l is the same for all the machines in precinct j.

Remark 1 *For candidate l in precinct j, for all voters $i = 1, \ldots, n_j$ and all machines $k = 1, \ldots, m_j$, suppose that (a) the probability of being assigned to a machine is the same for each voter ($\pi_{ijk} = \pi_{jk}$) and (b) the vote choice does not depend on the machine or on how other voters are assigned to machines ($\rho_{ijkl} = \rho_{ijl}$). Then the same vote share is expected for candidate l on every machine used to count votes in the precinct, i.e.,*

$$\text{for all } k, k' = 1, \ldots, m_j, \quad \tilde{R}_{jkl} = \tilde{R}_{jk'l} = n_j^{-1} \sum_{i=1}^{n_j} \rho_{ijl} . \tag{1}$$

If condition (1) holds, then the proportion of votes cast for candidate l on one machine in a precinct should not be systematically different from the proportion of votes cast for l on the other machines in the precinct. The proportion for l on the other machines should tend to be a good predictor for the proportion observed on machine k. When computing these predictor proportions I add small constants to both the numerator and denominator counts in cases where candidate l receives no votes on some set of $m_j - 1$ machines in a precinct, and I add small constants to the denominator counts in cases where candidate l receives all the votes on some set of $m_j - 1$ machines. These adjustments avoid making excessively sharp predictions. Formally, let n_{jk} denote the number of votes observed in precinct j on each machine k, with n_{jkl} denoting the number of votes on that machine for candidate l. Let $\delta_{kk'} = 1$ if $k = k'$, otherwise $\delta_{kk'} = 0$. Assuming $m_j > 1$, , define adjustment indicators z_{jl} and a_{jl}:

$$z_{jl} = \begin{cases} 1, & \text{if, for any } k = 1 \ldots m_j,\ \sum_{k'=1}^{m_j}(1-\delta_{kk'})n_{jk'l} = 0 \\ 0, & \text{otherwise,} \end{cases}$$

$$a_{jl} = \begin{cases} 1, & \text{if, for any } k = 1 \ldots m_j,\ \sum_{k'=1}^{m_j}(1-\delta_{kk'})n_{jk'l} = \sum_{k'=1}^{m_j}(1-\delta_{kk'})n_{jk'} \\ 0, & \text{otherwise.} \end{cases}$$

Assuming $m_j > 1$, the proportion of votes for l predicted for machine k, using the votes for l on the machines other than k in precinct j, is

$$\check{p}_{jkl} = \frac{\sum_{k'=1}^{m_j}(1-\delta_{kk'})(n_{jk'l} + z_{jl}/2)}{\sum_{k'=1}^{m_j}(1-\delta_{kk'})(n_{jk'} + z_{jl}/2 + a_{jl}/2)}, \tag{2}$$

The adjustment indicators cause the constant 1/2 to be added to all the counts for machines in a precinct if any machine in the precinct would otherwise be facing a predicted proportion of zero or one based on the votes recorded on the other machines in the precinct.

I use the Pearson chi-squared statistic to implement a randomization test of whether (1) holds for each precinct. For precinct j the test statistic is

$$X_{jl}^2 = \sum_{k=1}^{m_j} \frac{(n_{jkl} - n_{jk}\check{p}_{jkl})^2}{n_{jk}\check{p}_{jkl}} .$$

Remark 1's assumptions (a) and (b) imply that every distribution of the observed votes among each precinct's m_j machines is equally likely, subject to the constraint that the number of votes on each machine remains constant throughout the permutations of the votes. Hence we may test for (1) by checking whether the value of X^2_{jl} obtained using the observed data is large compared to the values obtained over all possible permutations of the observed votes. We fix the machine totals n_{jk} and the total number of votes for candidate l across all of the machines but shuffle the votes among the machines to obtain new sets of counts, say n^*_{jkl}. The constraint that the total number of votes for candidate l across all of the machines is fixed means that $\sum_{k=1}^{m_j} n^*_{jkl} = \sum_{k=1}^{m_j} n_{jkl}$. For each set of shuffled votes we compute the chi-squared statistic,

$$X^{2*}_{jl} = \sum_{k=1}^{m_j} \frac{(n^*_{jkl} - n_{jk}\check{p}^*_{jkl})^2}{n_{jk}\check{p}^*_{jkl}} ,$$

where $\check{p}^*_{jkl}$ denotes the predicted proportion (2) computed using the shuffled data. Because the number of permutations of the votes is large even for moderate numbers of votes and machines, I use a Monte Carlo approach that involves randomly sampling permutations in order to approximate the probabilities of observing values of X^{2*}_{jl} as large as X^2_{jl} or larger given the hypothesis that Remark 1's assumptions (a) and (b) hold. That is, assuming that (a) and (b) of Remark 1 hold, I estimate

$$g_{jl} = \text{Prob}\left(X^{2*}_{jl} \geq X^2_{jl} \mid m_j,\ \{n_{jk} : k = 1, \ldots, m_j\},\ \sum_{k=1}^{m_j} n_{jkl} \right) .$$

Let $\hat{g}_{jl}$ denote the Monte Carlo estimate of g_{jl}.

To combine the test results from the many precincts there are to assess from each county, I use the false discovery rate (FDR) (Benjamini and Hochberg 1995; Benjamini and Yekutieli 2005). The randomization method treats each precinct independently, so it is appropriate to use the form of the FDR that assumes independence. Benjamini and Hochberg (1995) define this FDR as follows. For candidate l, sort the values $\hat{g}_{jl}$ from all J precincts from smallest to largest. Let $\hat{g}_{(j)l}$ denote these ordered values, with $\hat{g}_{(1)l}$ being the smallest. For a chosen test level α (e.g., $\alpha = .05$), let d be the smallest value such that $\hat{g}_{(d+1)l} > (d+1)\alpha/J$. This number d is the number of tests *rejected* by the FDR criterion. If Remark 1's assumptions (a) and (b) hold for all machines in all precincts, then we should find $d = 0$.

A limitation of this method is that in precincts where all or all but one of the machines have very small counts n_{jkl} or $n_{jk} - n_{jkl}$, the number of distinct possible values of X^2_{jl} may be too small for the test based on the smallest observed tail probability to have any power. For instance, if $\alpha = .05$ and $J = 757$ (roughly the number of precincts in Miami-Dade County), then $\alpha/J \approx .000066$. A tail probability that small cannot occur in a precinct having three machines with n_{jk} values (1, 3, 1) and n_{jkl} values (1, 0, 0), as occurs in the ballot image data with the votes for president in one Miami-Dade election-day precinct. To mitigate this problem, I include in the analysis only precincts for which there are at least two machines k for which for candidate l we have both $\sum_{k'=1}^{m_j}(1-\delta_{kk'})n_{jk'l} > 1$ and $\sum_{k'=1}^{m_j}(1-\delta_{kk'})(n_{jk'} - n_{jk'l}) > 1$.[3]

Data

I apply the randomization test to voting data from the 2004 general election in three Florida counties: Broward, Miami-Dade and Pasco (see the Data Note for details on sources and contents of the data). Table 1 shows the number of precincts in each county. On election day, some machines were used to record votes from more than one precinct. This occurred in cases where more than one precinct shared a polling place. Most voting occurred on election day, November 2, 2004, but the data also include votes cast during the 15-day early voting period (October 18 through November 1, 2004). Table 1 also shows the number of early voting sites used in each county (earlyvoting.org 2004; Miami-Dade County 2004; Browning 2004). In Broward and Pasco counties, voters from all precincts could vote at any early voting site. In Miami-Dade county, voters from each precinct could vote only at selected early voting locations. At early voting sites each voting machine was used for voters from multiple precincts. The voting data for the early voting period do not directly indicate the voter's precinct but instead indicate which of several ballot styles the voter used. Table 1 shows the number of styles used during early voting for each county.

*** Table 1 about here ***

The randomization test is meaningful for precinct j only if at least in principle every voter is equally likely to use each of the machines. The realities of voting in the Florida counties present some challenges to this requirement.

The most obvious challenges concern early voting. For much the same reason that we separate

[3] Probably it would be better to include only precincts where there are at least $1/\alpha$ possible permutations of the votes for candidate l, subject to holding constant the machine totals $(n_{j1}, \ldots, n_{jm_j})$.

the election day votes cast in different precincts from one another, we would also like to avoid grouping together votes cast at different early voting sites. Voters using different sites probably live in different places and are likely to have significantly different preferences. Moreover, in Miami-Dade, not every ballot style was available on every voting machine at each early voting site, so not every voter could use every machine. Unfortunately, neither the ballot nor the event log files contained any indication of the physical location where each voting machine was used. I used Personal Electronic Ballot (PEB)[4] codes recorded in the event log files to group machines together, the idea being that machines for which the same PEB was used must have been located at the same early voting site.[5]

Another concern with early voting is that not every voting machine was used every day during the early voting period. I used the event log files to identify the dates during the early voting period when each voting machine was used. I grouped machines together only if they were used on all the same days. The "site-days" entries in Table 1 show the number of unique combinations of the PEB-based location groupings with these date goupings in Broward and Pasco counties, and the "style-site-days" entry shows the number of unique combinations of the PEB-based location and ballot style groupings with the date goupings in Miami-Dade County. These serve as the "precincts" j for the early voting randomization tests. The "site-day-machines" and "s-s-d-machines" entries show the number of unique combinations of the site-days or style-site-days goupings with voting machines. These are the "machines" k for the early voting randomization tests.

Much as machines being used on different days is a concern during the early voting period, there is also a potential problem due to machines being used at different times during each day. Figure 1 illustrates several patterns of potential concern. The plots in the figure show the times at which votes were cast on each voting machine on election day in four Miami-Dade precincts. Each row of letters in each plot indicates the time at which a "vote cast" transaction occurs for a voting machine in the event log files, with a letter being plotted at each point when a vote was recorded. There is one row of letters for each voting machine used in each precinct. Times are shown using

[4]For a description of how PEBs are used in Election Systems & Software "iVotronic" voting machines, see (Electronic Frontier Foundation 2004).

[5]For Miami-Dade County it was possible to supplement the PEB information with copies of files that showed the location of all but 88 of the machines used during early voting. See the Data Note for details.

a 24-hour clock and resolved to the second. In precinct 109, most of the machines were used throughout the day, but the machine labeled "e" was not used after 10am. A reasonable guess is that the machine was pulled from service at that time. In precinct 233, the machine labeled "c" was not used after 8am, and the machine labeled "f" was not used before 2pm. In precinct 322, the machine labeled "b" was used only between 11:30am and 2:30pm. In precinct 326, the machines labeled "g" and "m" were used only after 1pm. If some machines were not available for use during substantial parts of the day, then Remark 1's assumption (a) is not satisfied. Questions about this assumption also arise for other machines that exhibit irregular usage. For instance, in precinct 109 the machine labeled "k" was used much less often in the afternoon than in the morning, and in precinct 326 the machine labeled "p" was used heavily only after 6pm.

*** Figure 1 about here ***

Instead of trying to exclude machines for which usage during the day seems not to match the pattern of the other machines in a precinct, I construct a measure of how similar the patterns of time usage are for a precinct's machines and examine whether the measure is related to the tail probability estimates $\hat{g}_{jl}$. Let t_{jki} denote the time (in seconds) at which vote i was cast on machine k in precinct j. For each machine k in precinct j, I compute

$$\tau_{jk} = \frac{1 + (n_j - n_{jk})^{-2} \sum_{k'=1}^{m_j} (1-\delta_{kk'}) \sum_{i=1}^{n_{jk}} \sum_{i'=1}^{n_{jk'}} (1-\delta_{kk'}) |t_{jki} - t_{jk'i'}|}{1 + (n_{jk})^{-2} \sum_{i=1}^{n_{jk}} \sum_{h=1}^{n_{jk}} |t_{jki} - t_{jkh}|} .$$

The denominator measures the mean absolute difference among the times at which votes were cast on machine k, and the numerator measures the mean absolute difference between the times at which votes were cast on machine k and the times at which votes were cast on every other machine k' in precinct j. The ratio τ_{jk} achieves the lower bound of 1.0 if the mean absolute difference among the voting times on machine k is the same as the mean absolute difference between the voting times on k and the voting times on the other machines. The ratio increases as the voting times on machine k tend to differ on average more from the times on the other machines than they differ from one another. To compute a summary measure for each precinct j, I compute the geometric mean of the ratios τ_{jk}, namely,

$$\tau_j = \left(\prod_{k=1}^{m_j} \tau_{jk} \right)^{1/m_j} .$$

For the four precincts shown in Figure 1, τ_j has the values $\tau_{109} = 1.06$, $\tau_{233} = 1.09$, $\tau_{322} = 1.02$ and $\tau_{326} = 1.05$. The largest values for a machine in each of those precincts is $\max_k(\tau_{109k}) = 4.2$, $\max_k(\tau_{233k}) = 33.8$, $\max_k(\tau_{322k}) = 3.5$ and $\max_k(\tau_{326k}) = 1.8$.

We might expect $\hat{g}_{jl}$ to decrease as the dissimilarity between machines—measured by either τ_j or $\max_k(\tau_{jk})$—increases. A weakness of this approach is that because it not possible to tell which ballot image corresponds to which event log entry, it is not possible to customize the vote-time dissimilarity measure for each candidate. Over all the votes cast, however, we can be reasonably sure that the times recorded in the event log filess do correspond to the votes recorded in the ballot image files. Table 2 shows that for the most part the total counts of voting events and of ballot images are the same for each voting machine.

*** Table 2 about here ***

Randomization Test Results

I examine the votes cast for the Republican and Democratic candidates for president (George W. Bush and John F. Kerry) and for U.S. Senator (Mel Martinez and Betty Castor). I also examine the votes Yes or No for eight state consitutional amendments that appeared on the ballot in Florida in 2004. These amendments are described in Table 3. In all cases I consider the shares for each candidate or for each amendment voting option out of all ballots cast, including in the denominator ballots for which no vote choice was indicated for the referent office or amendment. I analyze the early voting data separately from the election day data.

*** Table 3 about here ***

Figure 2 shows a typical pattern for the distribution of the estimates $\hat{g}_{jl}$. The values depicted are for election day precincts in Miami-Dade County. Most of the values are much larger than the test level $\alpha = .05$.

*** Figure 2 about here ***

There is no tendency for $\hat{g}_{jl}$ to decrease as the dissimilarity in vote times between the machines in a precinct increases. The $\hat{g}_{jl}$ values are not significantly correlated across precincts with either τ_j or $\max_k(\tau_{jk})$. Indeed, for the Miami-Dade County election day data only seven of the twenty product moment correlations with each dissimilarity measure are negative, and the most negative value found is $\text{cor}(\hat{g}_{jl}, \max_k(\tau_{jk})) = -0.06$, for the Amendment 3 No votes.[6] Similar results are found for the correlations between $\log(\hat{g}_{jl})$ and $\log(\tau_j)$ and between $\log(\hat{g}_{jl})$ and $\log(\log(\tau_j))$.

The FDR test results reported in Tables 4, 5 and 6 do not provide much support for the idea that the votes cast in each precinct were randomly and independently assigned to each machine used in the precinct. For all three counties, in both the election day and the early voting data, there are many rejections of the hypothesis that (1) holds. There are somewhat more rejections among the election day vote counts. Pasco County early voting has the fewest rejections, with one rejection each for the Amendment 5 Yes votes and for the Amendment 7 No votes. For early voting in Broward County there are four rejections, for four of the amendment options. Notwithstanding the attempt to compare only similar machine counts to one another in the Miami-Dade County early voting data, by separating votes that occur at different sites, on different days and using different ballot styles, there are rejections in those data for nine of the twenty candidate and amendment options. The election day results show rejections for ten of the twenty options in Miami-Dade, thirteen of the twenty options in Broward and five of the twenty options in Pasco County.

*** Tables 4, 5 and 6 about here ***

On balance it seems unlikely that voting time dissimilarities between the machines in each precinct can explain the pattern of rejections for the election day votes. We have already reviewed the pattern of insignificant $\text{cor}(\hat{g}_{jl}, \tau_j)$ and $\text{cor}(\hat{g}_{jl}, \max_k(\tau_{jk}))$ values for the Miami-Dade election day data. For the Broward County data, only five of the $\text{cor}(\hat{g}_{jl}, \tau_j)$ values and only six of the $\text{cor}(\hat{g}_{jl}, \max_k(\tau_{jk}))$ values are negative, and all of those correlations are very small. The largest in magnitude is $\text{cor}(\hat{g}_{jl}, \max_k(\tau_{jk})) = -0.04$ for the Amendment 6 No vote. For the Pasco county data, six $\text{cor}(\hat{g}_{jl}, \tau_j)$ values and nine $\text{cor}(\hat{g}_{jl}, \max_k(\tau_{jk}))$ values are negative, but these correlations are again small. The largest in magnitude does occur for one of the FDR rejections, namely the Amendment 7 No votes, for which $\text{cor}(\hat{g}_{jl}, \tau_j) = -0.11$. But $\text{cor}(\hat{g}_{jl}, \tau_j) = -0.10$ for the Amendment 7 Yes votes, and for those votes there are no FDR rejections. For the votes for Bush and Kerry, which each show more than one FDR rejection, $\text{cor}(\hat{g}_{jl}, \tau_j) > 0$. For these latter two votes the $\text{cor}(\hat{g}_{jl}, \max_k(\tau_{jk}))$ values are negative but small, respectively -0.03 and -0.02. There is no significant relationship between the correlations $\text{cor}(\hat{g}_{jl}, \tau_j)$ or $\text{cor}(\hat{g}_{jl}, \max_k(\tau_{jk}))$ and the

[6]The largest positive correlation in the Miami-Dade data is 0.11 for τ_j, for the Amendment 5 No votes.

number of FDR rejections for a particular candidate or amendment option.[7]

Using Benford's Law to Test for Fradulent Votes

One method that has been suggested for testing whether reported vote totals are fraudulent is to compare the digits occurring in the vote counts to the distribution of digits expected under Benford's Law. Benford's Law specifies that the different digits should not occur with equal frequency. That is, each of the nine possible first significant digits $(1, 2, \ldots, 9)$ should not each occur one-ninth of the time, each of the ten possible second significant digits $(0, 1, \ldots, 9)$ should not each occur one-tenth of the time, and so forth. Instead, according to Benford's Law the first and second significant digits should occur with the frequencies shown in Table 7. Tests against Benford's Law have been promoted for use to detect fraud in forensic financial accounting (Durtschi, Hillison, and Pacini 2004). In the realm of vote count data the relevance of Benford's Law has been controversial. Pericchi and Torres (2004) use tests of the second digits of vote counts against the Benford's Law distribution to raise the prospect of fraud in the Venzuelan recall referendum of August 15, 2004. This charge was specifically denied in the Carter Center report (Carter Center 2005, 132–133), based on technical analysis reported in Brady (2005) and Taylor (2005).

*** Table 7 about here ***

Why should Benford's Law apply to vote count data? A fundamental result is that Benford's Law does not in general hold for data that are simply random (Raimi 1976; Hill 1995). This property is one basis for its proposed use in financial fraud detection. If someone uses numbers taken directly from a table of random numbers to fill out faked financial records, the digits will occur with equal frequency. The positive case for using Benford's Law with financial data is not altogether perspicuous, however. Durtschi et al. (2004), for instance, rely on the supposedly complicated origins of financial data as the rationale for expecting Benford's Law to hold:

"Boyle (1994) shows that data sets follow Benford's Law when the elements result

[7] In the Broward County data, a Poisson regression of the number of FDR rejections on the values of $\text{cor}(\log(\hat{g}_{jl}), \log(\max_k(\tau_{jk})))$ shows a marginally significant positive relationship: the coefficient estimate is 9.6 with a standard error of SE=5.8. But in the Miami-Dade County data the same kind of analysis shows a significant negative relationship between the same variables: the coefficient estimate is -18.1 (SE=4.9). In the Pasco County data the corresponding analysis produces a coefficient estimate of -1.9 (SE=6.0).

from random variables taken from divergent sources that have been multiplied, divided, or raised to integer powers. This helps explain why certain sets of accounting numbers often appear to closely follow a Benford distribution. Accounting numbers are often the result of a mathematical process. A simple example might be an account receivable which is a number of items sold (which comes from one distribution) multiplied by the price per item (coming from another distribution)." (Durtschi et al. 2004, 20–21)

The complexity rationale runs afoul of the way behavioral political scientists usually think about voting data. Students of voting behavior have developed a repertoire of models built on the idea that each individual's vote choice is essentially a coin flip (i.e., a stochastic choice). For some elections the coin many have more sides than two, and for different people the probabilities of the various outcomes are different. But the overall vote counts are seen as merely the sum of all the different coin flip outcomes. Such a sum of random coin flips lacks the complexity needed to produce the Benford's Law pattern in the vote counts' digits. Taking voter turnout decisions into account does not essentially change the basic coin flip idea. In this case, to produce the coin flip probabilities the probability that each person votes is multiplied by the conditional probability that the person makes a particular choice among the candidates or ballot initiative options.

One can see this standard behavioral perspective at work in the analysis used to support the conclusions reached about the Venezuelan referendum by the Carter Center. This is most explicit in the analysis reported by Taylor (2005). Taylor writes, "we use the multinomial model (4) of a 'fair election' and find that its significant digit distribution is virtually identical to the observed distribution, which is different than Benford's Law" (Taylor 2005, 22). Taylor also generates data using a Poisson model. As a general matter these two models are essentially the same—as Taylor (2005, 9) observes, the multinomial arises upon conditioning on the total of a set of Poissons. Neither has the complexity needed to produce digits that follow Benford's Law.

The kind of complexity that can produce counts with digits that follow Benford's Law refers to processes that are statistical mixtures (e.g., Janvresse and de la Rue (2004)), which means that random portions of the data come from different statistical distributions. There are some limits that apply to the extent of the mixing, however. If the number of distinct distributions is large, then the result is likely to be well approximated by some simple random process that does not satisfy Benford's Law. So if we are to believe that in general Benford's Law should be expected to describe the digits in vote counts, we need to have a behaviorally realistic process that involves mixing among a small number of distributions.

Another issue concerns whether Benford's Law should be expected to apply to all the digits in reported vote counts. In particular, for precinct-level data there are good reasons to doubt that the first digits of vote counts will satisfy Benford's Law. Brady (2005) develops a version of this argument. The basic point is that often precincts are designed to include roughly the same number of voters. If a candidate has roughly the same level of support in all the precincts, which means the candidate's share of the votes is roughly the same in all the precincts, then the vote counts will have the same first digit in all of the precincts. Imagine a situation where all precincts contain about 1,000 voters, and a candidate has the support of roughly fifty percent of the voters in every precinct. Then most of the precinct vote totals for the candidate will begin with the digits '4' or '5.' This result will hold no matter how mixed the processes may be that get the candidate to roughly fifty percent support in each precinct. For Benford's Law to be satisfied for the first digits of vote counts clearly depends on the occurrence of brittle accidents in the distribution of precinct sizes and in the alignment of precinct sizes with each candidate's support. It is difficult to see how there might be some connection to generally occurring political processes. So we may turn to the second significant digits of the vote counts, for which at least there is no similar knock down contrary argument.

For an example that illustrates these ideas, consider Table 8. This table reports Pearson chi-squared statistics for two kinds of tests. First is whether the distributions of the first digits of the precinct vote counts for the major party candidates for president and for U.S. Senator and for the eight constitutional amendments on election day 2004 in Miami-Dade County match the distribution specified by Benford's Law. Second is whether the first digits occur equally often. For the Benford's Law test, let $q_{B_1 i}$ denote the expected relative frequency with which the first significant digit is i. These $q_{B_1 i}$ values are the values shown in the first line of Table 7. Let d_{1i} be the number of times the first digit is i among the J precincts being considered, and set

$d_1 = \sum_{i=1}^{9} d_{1i}$. The statistic for the first-digit Benford's Law test is

$$X^2_{B_1} = \sum_{i=1}^{9} \frac{(d_{1i} - d_1 q_{B_1 i})^2}{d_1 q_{B_1 i}} \, .$$

For the test that first digits occur equally frequently, the test statistic is

$$X^2_{U_1} = \sum_{i=1}^{9} \frac{(d_{1i} - d_1/9)^2}{d_1/9} \, .$$

Assuming independence across precincts, these statistics may be compared to the χ^2-distribution with 8 degrees of freedom.[8] That distribution has a critical value of 15.5 for a .05-level test. Since all of the statistics reported in Table 8 greatly exceed that value, the hypothesis that the first significant digits follow Benford's Law may be handily rejected, as may be the hypothesis that the nine values (1–9) occur equally often.

*** Table 8 about here ***

In contrast, consider Table 9, which reports Pearson chi-squared statistics for tests of the distribution of the vote counts' second significant digits. For $q_{B_2 i}$ denoting the expected relative frequency with which the second significant digit is i (given by the second line in Table 7), and with d_{2i} being the number of times the second digit is i among the J precincts being considered and $d_2 = \sum_{i=0}^{9} d_{2i}$, the statistic for the second-digit Benford's Law test is

$$X^2_{B_2} = \sum_{i=0}^{9} \frac{(d_{2i} - d_2 q_{B_2 i})^2}{d_2 q_{B_2 i}} \, .$$

For the test that second digits occur equally frequently, the test statistic is

$$X^2_{U_2} = \sum_{i=0}^{9} \frac{(d_{2i} - d_2/10)^2}{d_2/10} \, .$$

These statistics may be compared to the χ^2-distribution with 9 degrees of freedom, which has a critical value of 16.9 for a .05-level test. The results, reported in the first two columns of Table 9, give little reason to doubt that Benford's Law applies. Two of the twenty statistics are larger

[8]The consequences of dependence are unclear. It may develop that calibration is necessary to establish the correct distribution, especially when the number of precincts is not large. Similar comments apply to the $X^2_{B_2}$ and $X^2_{U_2}$ statistics introduced for second digits below.

than the critical value for a .05-level test. But if we consider the twenty tests to be independent, then with a single-test level of $\alpha = .05$, using the FDR gives no reason to be concerned unless we obtain a statistic larger than 25.46 (with a single-test level of .10, using the FDR establishes a 23.59 as the value beyond which we should be concerned).[9] The largest $X^2_{B_2}$ value in the first column of Table 9 is 17.9. The results give reason to reject the assumption that the second digits are equally likely to take any of the ten possible values. The largest $X^2_{U_2}$ value in the second column of Table 9 is 25.3.

*** Table 9 about here ***

The remaining columns of Table 9 show that what works for precincts need not work for voting machines. The middle columns report the results of applying the tests to the vote counts on the individual voting machines used on election day in Miami-Dade County. Acknowledging that some voting machines in Miami-Dade recorded votes from more than one precinct on election day, the last two columns show results from applying the tests to vote counts for each unique precinct-machine combination. Both forms of the analysis firmly reject the idea that Benford's Law describes the distribution of the second significant digits of the vote counts on election day voting machines in Miami-Dade County.

Generating Vote Counts that Satisfy Benford's Law

Is there a family of processes that are behaviorally plausible from a political point of view and that are capable of producing precinct-level vote counts that satisfy Benford's Law for the second significant digits but not for the first significant digits? Can we explain why such a process would produce precinct counts that satisfy the second-digit Benford's Law but not machine counts that do so?

The second question has an answer that does not depend on the details of how the precinct counts may be generated, so let's consider it first. The point is to remember that a random process that is not a mixture does not in general produce digits that satisfy Benford's Law. Using that fact, we can explain the non-Benford machine counts in cases where votes are randomly assigned to the voting machines being used in each precinct. If the probability that each vote cast

[9] For 20 independent tests and single-test level $\alpha = .05$, the FDR gives $0.0025 = .05/20$ as the first tail probability to be concerned about, which for the χ^2-distribution with 9 degrees of freedom corrsponds to a critical value of 25.46. The value of 23.59 is obtained analogously.

in precinct j is assigned to machine k is π_{jk}, then conditioning on the total number of votes cast in each precinct, the distribution of votes among the machines in precinct j is multinomial with outcomes proportional to $\pi_j = (\pi_{j1}, \ldots, \pi_{jn_j})$. If the probability vectors π_j or the total number of votes cast vary across precincts, these multinomial distributions may vary considerably from precinct to precinct, but having a collection of vectors of counts each generated by a different multinomial distribution does not in general give counts that satisfy Benford's Law.

So what can produce precinct-level vote counts that satisfy the second-digit Benford's Law? For a behaviorally realistic process that involves mixing among a small number of distributions, we can think about political parties, or more generally about the coalitions that come together at election time. Usually each candidate (or each side) has a collection of core supporters. These core supporters are virtually certain to vote for their side. Viewed as coins, we might say these core supporters always come up "heads." Note that this virtual certainty of support for one candidate need not imply any loyalty to the candidate that lasts longer than election day. But at the time the candidate votes, it is there. Any voter who is not such a core supporter for any side may possibly vote for any of the available alternatives.[10] Using the mean probability that such available voters vote for each candidate, we obtain a model where the total vote for a candidate in each precinct is a mixture of two distributions: the distribution of core supporters and the distribution of available voters.

The following **R** (R Development Core Team 2003) function generates vote counts for one candidate across a set of simulated precincts from such a model.

```
pbenf <- function(size, nprecincts=500, lsplit=.1, hsplit=.1, bfrac=1/2) {
  z <- sapply(1:nprecincts,
   function(x){
     p2 <- c(runif(1,0,lsplit),runif(1,(1-hsplit),1));
     pf <- c(rbeta(1,1,bfrac),rbeta(1,bfrac,1));
     partypm <- rpois(2,size*pf/sum(pf));
     sum(votes <- rpois(length(partypm), lambda=partypm*p2))
   })
}
```

For each of the `nprecincts` simulated precincts the vector `p2` contains two numbers. The first

[10] I think it may be better to distinguish between those voters who have firmly made up their minds for whom they will vote when they arrive at the polls and those who have not. This would give a distinction between, say, "committed" and "undecided" voters. In future drafts of this paper I will likely shift to something like that usage.

number, drawn uniformly from the interval $[0, \texttt{lsplit}]$, represents the probability that available voters vote for the candidate. The second number, drawn uniformly from the interval $[1 - \texttt{hsplit}, 1]$, represents the probability that the candidate's core voters vote for the candidate. The vector `pf` represents the proportion of the voters in each precinct who are expected to be of each type. With the default argument value $\texttt{bfrac} = 1/2$, the first, Beta-distributed value in `pf` has a mean of 2/3 and the second value has a mean of 1/3. The vector `partypm` contains the Poisson-distributed expected number of voters of each type. The vector `votes` contains the vote counts for the candidate from each type of voters in each precinct. These are summed to give the overall number of votes for the candidate in each precinct.[11]

Tables 10 and 11 show the results of a Monte Carlo simulation exercise using function `pbenf` to generate precinct vote counts for various choices of the function's arguments. The parameter denoted Size in the table refers to the `size` argument, which is the expected number of voters in each precinct. All the precincts generated by one invocation of `pbenf` have the same expected number of voters, although the actual number, which is Poisson distributed, varies over precincts. The parameter denoted Split in the table refers to the `lsplit` argument (the `hsplit` argument always has the value 0.1). The values in the Mean Votes column indicate the number of votes the candidate is expected to receive in each precinct given the corresponding parameter values.[12]

*** Tables 10 and 11 about here ***

In Table 10 one can see that in most cases the simulated vote counts satisfy the second-digit Benford's Law. In Table 11 the simulated vote counts satisfy the second-digit Benford's Law for small values of `lsplit` and Size values up to about Size=2000, and for for larger values of `lsplit` and Size=3000, but mostly not for Size values 2250, 2500 and 2750. These results suggest that the electoral coalition model that features two types of voters for each candidate can generate vote counts with second digits that satisfy Benford's Law for a wide variety of parametric

[11]If the only goal is to produce counts whose second digits usually satisfy the second-digit Benford's Law, then it is not necessary to have the expected number of voters (`partypm`) and the vote counts (`votes`) be Poisson distributed. If the `pbenf` function is changed to use the assignments `partypm <- size*pf/sum(pf)` and `votes <- partypm*p2`, then we get second-digit Benford's Law results very similar to those obtained for the baseline model for the conditions considered in the Monte Carlo simulations reported in Tables 10 and 11. This alternative specification demonstrates that the essential feature that produces the second-digit Benford's Law pattern is the mixture of the core and available voting groups, not variation that may occur in the sizes of voting precincts. Using the Poisson-distributed values may impart greater realism, and it is noteworthy that doing so does not reduce the function's ability to produce counts with digits that satisfy the second-digit Benford's Law.

[12]The `bfrac` argument always equals the default value, $\texttt{bfrac} = 1/2$.

configurations, although clearly not for all possible parameter values. Hence the electoral coalition model (or improved versions of it) may possibly explain the patterns we see in real election data. By the way, the vote counts produced by the `pbenf` function do not have first significant digits that satisfy the first-digit Benford's Law.

Can Benford's Law Detect Vote Fraud?

Applying the second-digit Benford's Law test to other vote count data from the 2004 election in Florida produces some results that suggest that Benford's Law applies to the data and other results that raise questions. Table 12 reports results based on data from early voting in Miami-Dade county. Applying the FDR of Benjamini and Hochberg (1995) to the twenty tests for site-style-days, the results look fine if we use a single-test level of $\alpha = .05$, since no $X^2_{B_2}$ value is greater than 25.46, but the results are problematic if $\alpha = .10$ ($X^2_{B_2}$ for the Amendment 7 Yes votes is 24.6, which is greater than 23.6). The election day precinct results for Broward, shown in Table 13, are similar. They are fine using the FDR with $\alpha = .05$ but problematic using $\alpha = .10$: two of the Amendment vote counts have $X^2_{B_2} > 23.6$. The Broward early voting results for counts at the level of ballot styles are fine if the FDR is used. The largest $X^2_{B_2}$ value among these early voting tests is $X^2_{B_2} = 21.4$, for the votes for Kerry. The election day results for Pasco, shown in Table 14, have one value of $X^2_{B_2}$ large enough to reject the hypothesis that Benford's law applies even using the FDR among the twenty tests with $\alpha = .05$. This is the value $X^2_{B_2} = 29.5$, which occurs for the Amendment 7 Yes votes. Considered on their own and using the FDR for twenty tests, the early voting machine-precinct results for Pasco are fine.

*** Tables 12, 13 and 14 about here ***

The results for voting machines in Tables 12, 13 and 14 further illustrate that the second-digit Benford's Law property mostly does not apply to the vote counts on machines in these Florida counties. The case that comes closest to being an exception is the machine results for early voting in Broward County. Many of those $X^2_{B_2}$ values are unproblematically small, but three are larger than the χ^2_9 critical value for a single test at level $\alpha = .05$, and two are large even when we use the FDR. For the Amendment 8 Yes votes we have $X^2_{B_2} = 27.9$, which is larger than the critical value for the FDR for twenty tests with $\alpha = .05$, and for the Amendment 7 Yes votes we have $X^2_{B_2} = 44.0$, which is very large by any standard.

The value $X^2_{B_2} = 29.5$ that occurs for the election day precinct data from Pasco County is large enough to count as a rejection of the second-digit Benford's Law hypothesis even using the FDR among all 60 of the election day tests, pooling across the three counties: the quantile of χ^2_9 corresponding to a tail probability of .05/60 is 28.35. If we pool over all 120 of the election day precinct and early voting site-style-day, style and machine-precinct tests, the value $X^2_{B_2} = 29.5$ is not problematic according to the FDR with $\alpha = .05$, since the quantile of χ^2_9 corresponding to a tail probability of .05/120 is 30.13. But using $\alpha = .10$ we again have a problem even when pooling over all 120 tests, because using the FDR we again arrive at the χ^2_9 quantile of 28.35.

Do the relatively large $X^2_{B_2}$ values for the precinct-level vote counts suggest the counts have been fraudulently manipulated? The simulations reported in Tables 10 and 11 suggest that an electorally intelligible and benign process can produce counts that often satisfy the second-digit Benford's Law. Suppose we take a process that we know usually produces such counts and perturb it in ways that mimic some ways vote fraud may occur. Does the Benford's Law test signal that there has been a distortion? If so, we might conclude that the relatively large $X^2_{B_2}$ values suggest that maybe there has been fraud. Because we know the Benford's Law test can fail even when there is nothing like fraud in the data generating process, such a result can do no more than suggest the possibility of fraud. But if the Benford's Law test does not catch perturbations that we inject into otherwise pristine data, then of course the test is not useful for detecting vote fraud. In this case the mostly clean precinct-level results should not give us any comfort.

I simulate three variations of each of two kinds of vote manipulation. The two basic manipulations I describe as (1) adding repeaters and (2) proportionally increasing or decreasing vote totals. The variations apply each manipulation either to all precincts or to precincts in which the unmanipulated votes fall above or below specified thresholds.

My conception of repeaters harks back to the classic manipulation Gumbel (2005) describes as having been perfected by several American city political machines in the late nineteenth and early twentieth centuries. Repeaters in the nineteenth century's Tammany Hall were the primary referents of the familiar phrase, "vote early and often." As Gumbel writes, "The repeaters carried changes of clothing, including several sets of coats and hats, so they could plausibly come forward a second or third or fourth time in the guise of an entirely new person.... Many of the repeaters sported full beards at the beginning of the day, only to end it clean-shaven" (Gumbel 2005, 74).

Nowadays repeaters might simply be a few lines of computer code hidden in a PEB.

I implement repeaters by adding to a candidate's vote total a number equal to a specified fraction of the expected number of voters in each precinct. The number of votes added does not depend on the number of votes the candidate would otherwise receive, so the number added is not a function of the candidate's true support. To implement this idea, I replace the last line in the function that is applied to each precinct in the `pbenf` function with the following two lines,

```
votes <- sum(rpois(length(partypm), lambda=partypm*p2))
votes + sum(partypm)*frac;
```

The argument `frac` specifies the fraction of the expected voter number that is to be added.

The idea of proportionally increasing or decreasing vote totals is intended to represent two kinds of situations. One is where votes from a candidate are simply tossed out. A proportional decrease in a candidate's votes corresponds to the case where a fixed proportion of the candidate's votes are discarded in each precinct. The other situation is where votes are swapped from one candidate to another candidate. The candidate from whom the votes are taken could suffer proportional decreases, while the candidate who is receiving the votes is experiencing proportional increases. It may be that the Benford's Law tests can detect either the decreases or the increases, but not both. I implement this idea by replacing the last line in the function that is applied to each precinct in the `pbenf` function with the following two lines,

```
votes <- sum(rpois(length(partypm), lambda=partypm*p2))
votes <- ceiling(votes*frac);
```

The argument `frac` specifies the proportion by which the votes are to be increased or decreased. There are increases if `frac` > 1 and decreases if `frac` < 1.

I also consider variations of repeaters and proportional adjustments in which the manipulations are done only for a subset of the precincts. The subset to which the manipulations are applied depends on the votes the candidate is receiving before the manipulation is applied. The threshold for applying the changes is always the number of votes the candidate is expected to receive in each precinct. For the simulation function `pbenf`, that expectation may be computed using the **R** code

```
meanpbenf <-
```

```
size*(1/(1+bfrac))*(lsplit/2) + size*(bfrac/(bfrac+1))*(1+hsplit)/2;
```

The "Mean Votes" column in Tables 10 and 11 reports these expected vote values for a number of combinations of parameter values. In the case I designate as "below threshold," the manipulation is applied if the candidate is receiving fewer than `meanpbenf` votes. In the "above threshold" case the manipulation is applied if the candidate is receiving more than `meanpbenf` votes.

In each case I simulate these vote manipulations starting with vote counts produced by `pbenf`, using parameters that tended to produce counts that satisfied the second-digit Benford's Law for a wide range of expected numbers of voters in each precinct. In particular, referring to Tables 10 and 11, I use Split = 0.1 (which is `lsplit` = .1). Using that Split value produced small values of $X^2_{B_2}$ for expected numbers of voters per precinct (i.e., "Size") ranging from 500 to 2,000 and precincts numbering from 500 to 1,000. Over that range of sizes, the Monte-Carlo estimated expected value of $X^2_{B_2}$ is always smaller than the expected value of $X^2_{U_2}$, and often the expected value of $X^2_{U_2}$ is very large.

The results in Table 15 show that the second-digit Benford's Law test can sometimes but not always detect distortions from repeaters acting the same way in all precincts. The column labeled Add in the table shows the value of `frac`, which indicates how many votes were added as a fraction of the expected number of voters in each precinct. For example, with Size = 500 and Add = 0.05, 25 votes were added to the candidates vote total in each precinct. We ask whether each averaged $X^2_{B_2}$ statistic shown in the table exceeds the critical value for χ^2_9 for a test at level $\alpha = .05$, which is 16.9. For Size = 2000 and 1,000 precincts, the average $X^2_{B_2}$ value is always larger than 16.9, which suggests the Benford's Law test would usually detect the manipulation in such precincts. With Size = 2000 and 500 precincts, the average $X^2_{B_2}$ is greater than 16.9 only for Add = 0.10 or larger. So in such precincts it appears the test would usually detect repeaters only if they were as numerous as ten percent of the bona fide voters. With Size = 1500, the test typically triggers only for Add greater than 0.20. With Size = 1000 or 500, the test triggers irregularly for some of the larger values of Add.

*** Table 15 about here ***

The results in Table 16 show that the Benford's Law test is somewhat better able to signal manipulation when the repeater manipulation occurs in the precincts where the candidate is otherwise getting more votes than would be expected based on the uncontaminated process, but the test does not do as well when the repeater manipulation is happening in precincts where the candidate is otherwise receiving fewer votes than would be expected. The averaged $X^2_{B_2}$ values shown in the Above Threshold columns are typically larger than the corresponding columns in Table 15, while the averaged $X^2_{B_2}$ values shown in the Below Threshold columns are typically smaller.

*** Table 16 about here ***

The results in Table 17 suggest that the Benford's Law test has only very limited ability to detect proportional increases or decreases in a candidate's vote that happen throughout all precincts. The "Prop." values in the table indicate the value of `frac` that was used. The values used range from a twenty percent reduction in the candidate's vote (Prop. = 0.8) to a twenty percent increase (Prop. = 1.2). The only situations in which significantly large average values of $X^2_{B_2}$ occur are for 1,000 precincts with Size = 2000 and Prop. equal to 1.1 or greater, or with Size = 500 and Prop. = 0.8. Since a proportional adjustment that affects all precincts the same way is indistinguishable from a candidate's simply receiving greater or lesser support throughout the electorate, it is perhaps not surprising that the Benford's Law test has little ability to detect such a manipulation.

*** Table 17 about here ***

The results in Table 18 show that the Benford's Law test is much more effective when there are proportional increases that occur in the precincts where the candidate is otherwise getting more votes than would be expected based on the uncontaminated process. With 1,000 precincts, the average $X^2_{B_2}$ values are significanly large in three-quarters of the Above Threshold instances where Prop. is greater than 1. With 500 precincts the average $X^2_{B_2}$ values are significantly large when Prop. is greater than 1 only for Size = 2000, with one exceptional case occurring for Size = 500 and Prop. = 1.15. The Benford's Law test is mostly not more effective at detecting the proportional adjustment manipulation when it is happening in precincts where the candidate is otherwise receiving fewer votes than would be expected based on the uncontaminated process. There are significantly large average $X^2_{B_2}$ values in the Below Threshold columns with Size = 500 and Prop. > 1, but for the most part the average $X^2_{B_2}$ values in the Below Threshold columns are not large.

*** Table 18 about here ***

While the Benford's Law test can detect proportional increases in a candidate's support in many situations where only some of the precincts are being affected, it is not very effective at detecting proportional reductions. In Table 18, the average $X^2_{B_2}$ values for most of the instances where Prop. < 1 are not large.

Benford's Law and Voting Machine Vote Counts

Whatever we may conclude about the extent to which the second-digit Benford's Law distribution applies to the precinct-level vote counts from the three Florida counties in 2004, the results in Table 9 and in the other tables show that the Benford's Law distribution in general does not apply to the vote counts on voting machines in these counties. Notwithstanding the evidence from the randomization tests that there is not much support for the idea that the votes cast in each precinct were randomly and independently assigned to the machines used in the precinct, I conjectured that random assignment of votes to machines may explain the non-Benford machine counts. Ignoring for a moment the question of how votes actually were assigned to machines in the counties, I now consider whether a process that does assign the votes randomly and independently does produce second-digit distributions that do not match the second-digit Benford's Law.

First I consider a process that has precincts that contain the same number of voters as were in the Miami-Dade election day precincts, but has votes determined according to mixture processes like those simulated in Table 10. To implement such a process in **R**, I create a matrix, `precinct.data`, that has two rows and as many columns as there are election day precincts. The first row contains the number of votes cast on election day in each precinct, and the second row contains the number of voting machines used on election day to record votes for that precinct.[13] The **R** function that uses the `precinct.data` matrix to simulate randomly assigning votes to machines is defined as follows.

```
pbenfm <- function(lsplit=.1, hsplit=.1, bfrac=1/2) {
  z <- apply(precinct.data, 2,
   function(x){
     p2 <- c(runif(1,0,lsplit),runif(1,(1-hsplit),1));
```

[13]The total number of machines referenced in the `precinct.data` matrix corresponds to the number of precinct-machines indicated in Table 1.

```
    pf <- c(rbeta(1,1,bfrac),rbeta(1,bfrac,1));
    size <- x[1];
    partypm <- rpois(2,size*pf/sum(pf));
    votes <- sum(rpois(length(partypm), lambda=partypm*p2))
    nmachines <- x[2];
    mach <- rep(0,nmachines);
    # allocate votes at random to the nmachines machines
    if (votes > 0) mach <- table(sample(1:nmachines, votes, replace=TRUE));
    return( mach )
  })
```

The `pbenfm` function does not constrain the total number of votes on each machine to correspond to the number actually recorded on the machine in the original election day data. In `pbenfm`, each vote is equally likely to be counted on each of each precinct's machines.

Running such a simulation with parameters taken from the previously reported simulations sometimes but not always produces a pattern matching what occurs in the actual data.[14] Results are reported in Table 19. For the chosen set of Split values, ranging from 0.1 to 0.7, the second-digit Benford's Law always describes the digits in the simulated precinct vote counts. For Split values larger than 0.4, the digits in the simulated machine counts do not follow the Benford's Law distribution, which matches the pattern in the original data. But for Split = 0.3 or smaller, the machine counts do satisfy Benford's Law. Random assignment of votes to machines does not necessarily annihilate the Benford's Law pattern.

*** Table 19 about here ***

Randomly assigning the votes actually cast on election day in Miami-Dade County comes close to reproducing the Benford's Law test results reported, for precinct-machines, in Table 9. The first row in Table 20 shows what happens if the votes cast for Bush and Kerry are randomly assigned to machines, using the same procedure as in `pbenfm`. That is, in that program, instead of using `votes` simulated using the statistical mixture process, the results for "actual precincts" in Table 20 use the original vote counts for the respective candidates. So the results for precincts in that row are simply taken from Table 9. For both Bush and Kerry, randomly assigning the votes produces average $X^2_{B_2}$ values that are only slightly smaller than the ones computed for the original precinct-machine counts. For the vote counts that actually occurred on election day, it seems that the approximation to random assignment to machines that did happen then is a large

[14]Parameters `hsplit` and `bfrac` are left at their default values.

part of the reason the machine vote counts are non-Benford.

*** Table 20 about here ***

Randomly assigning vote counts produced by simulations calibrated to mimic the votes actually cast on election day in Miami-Dade County muddies the waters a bit. Such results are reported in the second and third lines in Table 20. To produce those simulations, I used `rgenoud` (Mebane and Sekhon 2005; Sekhon and Mebane 1998) to find values for the parameters of the version of `pbenf` (using the Miami-Dade precinct sizes) that minimize the discrepancy between the second digits of the votes expected for each candidate and the second digits of the actual vote counts. Specifically, I used `meanpbenf` with `size` set equal to the actual Miami-Dade election day precinct sizes to compute expected vote counts, then chose values for the `lsplit` and `hsplit` parameters to minimize a chi-squared statistic in which the distribution of the digits of the expected vote counts produced by `meanpbenf` provides the expected values. Results using this calibration appear in the second line of Table 20.[15] The results in the third line of Table 20 follow upon using a version of the vote simulating function in which four parameters are calibrated. The expected vote function in this case is the following

```
meanpbenfB <-
  size*(1/(1+lbfrac))*(lsplit/2) + size*(hbfrac/(hbfrac+1))*(1+hsplit)/2
```

With `meanpbenfB` I used `rgenoud` to minimize discrepancies with both the second digits of the counts and the counts themselves.[16] Figure 3, which presents density plots to compare the calibrated simulations to the actual precinct vote counts, suggests the calibrated simulations provide a better fit to the votes for Bush than to the votes for Kerry. In any case, neither the two-parameter calibration nor the four-parameter calibration leads to machine vote counts that consistently deviate from the second-digit Benford's Law distribution.[17]

*** Figure 3 about here ***

All told, nearly random assignment of votes to voting machines may explain the non-Benford machine counts so frequently observed in the data from the three counties, but it is not

[15] The calibration values for Bush are `lsplit` = 0.1168443, `hsplit` = 0.5699924. For Kerry the values are `lsplit` = 0.1789472, `hsplit` = 0.6468790.

[16] The calibration values for Bush are `lsplit` = 0.1144489, `hsplit` = 0.9947601, `lbfrac` = 3.0359998, `rbfrac` = 2.6032223. For Kerry the values are `lsplit` = 0.4803455, `hsplit` = 0.9807219, `lbfrac` = 0.2774467, `rbfrac` = 2.1231147.

[17] The calibrated simulation results presented in Table 20 use fixed precinct voter sizes; i.e., they use `partypm <- size*pf/sum(pf)` and `votes <- partypm*p2`.

appropriate to draw from that any wider message about how such randomization may affect Benford's Law tests. It is not clear what may be true in general.

Discussion

Both the vote randomization test and the second-digit Benford's Law appear potentially useful for detecting election fraud. In both cases a number of issues remain unsettled.

The vote randomization test finds strong evidence that votes were not randomly and independently assigned to the various voting machines in use in precincts on election day in the three Florida counties. The test also suggests that votes were not randomly distributed among comparable machines during the early voting period. The principle question is why do the candidate and amendment option vote shares differ across machines. One innocent possibility is that we have not successfully grouped the machines into comparable sets. Differences in usage times during each day may explain the different vote shares. The measure τ_{jk} may not be adequate, or my use of it may not be correct. There is also at least one distinction among voting machines that is not reflected in the tests reported in this paper. Some machines were specially equipped with audio capability to support independent voting by visually impaired voters. Perhaps the voters who used such machines had distinctive preferences. I did not separate out the audio-enabled machines principally because information to identify them all is lacking. I have information that identifies some of the audio-enabled machines in Miami-Dade County, but even for the machines designated as audio-enabled it is not clear from the records I have whether the audio capabilities were operating while the machines were being used.

Three classes of questions remain regarding the Benford's Law tests. First, this paper only suggests the range of mixture processes that might be behaviorally defensible and also tend to produce counts with digits that satisfy Benford's Law. Can processes with more heterogeneity in each precinct work? The simulations I have conducted so far to explore that suggest the situation is complicated. Second, how can we make sense of the fact that the mixture process produces counts that satisfy the second-digit Benford's Law for many but not all combinations of parameters? Third, what parameter values produce counts that closely match the counts that occur in real elections? The small calibration effort I attempted produced a pretty good approximation to the counts for Bush on election day in Miami-Dade County but did not do as well for the counts for Kerry. Can calibration be elevated to become proper estimation? For instance, is there a rationale for treating the second digits of a set of counts as if they were sufficient statistics?

Data Note

David Dill supplied ballot and event log files recovered from electronic voting machines in Broward, Miami-Dade and Pasco counties. The files were originally obtained by Martha Mahoney using open records requests funded by the Verified Voting Foundation. The ballot files indicate the choices made for each office by each voter and include labels identifying for each ballot the voting machine and the precinct (for election day ballots) or ballot style (for early voting ballots). The event log files show the time (resolved to the second) at which various transactions occurred on each machine, including the time at which each vote was recorded. It is not possible to match vote choices in the ballot files to voting events in the event log files.

Early voting polling site locations for many of the Miami-Dade machines was taken from a file supplied by Martha Mahoney (file "ev.xls," received by me on August 16, 2005). Of the 670 machines that recorded votes during early voting in Miami-Dade, 88 are not included in that file. Two files supplied by Martha Mahoney also were used to determine which Miami-Dade machines were operating with audio capability enabled. These are the "ev.xls" file and a file "Election.xls" (received by me on August 16, 2005) for the machines used on election day.

The data comprise files for electronic early voting and electronic polling place votes but do not include information about paper absentee votes.

References

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B* 57 (1): 289–300.

Benjamini, Yoav and Daniel Yekutieli. 2005. "False Discovery Rate-Adjusted Multiple Confidence Intervals for Selected Parameters." *Journal of the American Statistical Association* 100 (Mar.): 71–81.

Brady, Henry E. 2005. "Comments on Benford s Law and the Venezuelan Election." MS dated January 19, 2005.

Browning, Kurt S. 2004. "Vote Notes." Quarterly Publication of the Pasco County Supervisor of Elections.

Carter Center. 2005. "Observing the Venezuela Presidential Recall Referendum: Comprehensive Report.".

Durtschi, Cindy, William Hillison, and Carl Pacini. 2004. "The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data." *Journal of Forensic Accounting* 5: 17–34.

earlyvoting.org. 2004. "Early Voting for Broward Voters in the November 2, 2004 General Election." URL http://earlyvoting.org/earlyvoting14sites.pdf (accessed January 20, 2006).

Electronic Frontier Foundation. 2004. "Electronic Voting Machine Information Sheet: Election Systems & Software iVotronic." URL http://www.eff.org/Activism/E-voting/20040818_ess_ivotronic_v0.8.pdf (accessed March 27, 2006).

Gronke, Paul, Benjamin Bishin, Daniel Stevens, and Eva Galanes-Rosenbaum. 2005. "Early Voting in Florida, 2004." Paper prepared for the Annual Meeting of the American Political Science Association, Washington, DC, September 1, 2005.

Gumbel, Andrew. 2005. *Steal This Vote*. New York: Nation Books.

Hill, Theodore P. 1995. "A Statistical Derivation of the Significant-digit Law." *Statistical Science* 10: 354–363.

Janvresse, Élise and Thierry de la Rue. 2004. "From Uniform Distributions to Benford's Law." *Journal of Applied Probability* 41: 1203–1210.

Lehoucq, Fabrice. 2003. "Electoral Fraud: Causes, Types, and Consequences." *Annual Review of Political Science* 6 (June): 233–256.

Mebane, Walter R., Jr., Jasjeet S. Sekhon, and Jonathan Wand. 2001. "Detection of Multinomial Voting Irregularities." *2001 Proceedings of the American Statistical Association*. Social Statistics Section [CD-ROM].

Mebane, Walter R., Jr. and Jasjeet Singh Sekhon. 2005. "rgenoud: R version of GENetic Optimization Using Derivatives.". Computer software: Comprehensive R Archive Network.

Miami-Dade County. 2004. "New, Expanded Hours Announced for Early Voting in Hialeah as Miami-Dade County Sets New Early Voting Record." News release, October 21, 2004.

Pericchi, Luis Raúl and David Torres. 2004. "La Ley de Newcomb-Benford y sus aplicaciones al Referendum Revocatorio en Venezuela." Reporte Técnico no-definitivo 2a. versión: Octubre 01,2004.

R Development Core Team. 2003. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-00-3.
URL `http://www.R-project.org`

Raimi, Ralph A. 1976. "The First Digit Problem." *American Mathematical Monthly* 83 (7): 521–538.

Sekhon, Jasjeet Singh and Walter R. Mebane, Jr. 1998. "Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models." *Political Analysis* 7: 189–203.

Taylor, Jonathan. 2005. "Too Many Ties? An Empirical Analysis of the Venezuelan Referendum Counts." MS dated November 7, 2005.

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

Table 1: Precinct, Machine and Ballot Statistics

Election Day	Broward	Miami-Dade	Pasco
Precincts	775	757	152
Machines	5,306	5,323	1,338
Precinct-machines	10,614	14,128	2,676
Ballots	431,488	435,902	127,526

Early Voting	Broward	Miami-Dade	Pasco
Sites	20	14	3
Styles	150	100	16
Site-days	110	—	4
Style-site-days	—	4,429	—
Machines	190	726	36
Site-day-machines	380	—	72
S-s-d-machines	—	24,374	—
Ballots	176,743	242,344	29,584

Table 2: Event Transaction Counts and Ballot Counts

	Early Voting			Election Day		
	Excess Ballots	Counts Match	Excess Events	Excess Ballots	Counts Match	Excess Events
Broward	0	190	0	15	5,290	1
Miami-Dade	2	724	0	14	5,309	0
Pasco	0	36	0	0	1,338	0

Note: Entries show the number of voting machines having each described relationship between the number of “Normal ballot cast” or “Super ballot cast” events in the event log files and the number of ballots in the ballot image files.

Table 3: Florida Constitutional Amendments on the Ballot in 2004

		Yes	No
Am. 1	Parental Notification of a Minor's Termination of Pregnancy	4,639,635	2,534,910
Am. 2	Constitutional Amendments Proposed by Initiative	4,574,361	2,109,013
Am. 3	The Medical Liability Claimant's Compensation Amendment	4,583,164	2,622,143
Am. 4	Authorizes Miami-Dade and Broward County Voters to Approve Slot Machines in Parimutuel Facilities	3,631,261	3,512,181
Am. 5	Florida Minimum Wage Amendment	5,198,514	2,097,151
Am. 6	Repeal of High Speed Rail Amendment	4,519,423	2,573,280
Am. 7	Patients' Right to Know About Adverse Medical Incidents	5,849,125	1,358,183
Am. 8	Public Protection from Repeated Medical Malpractice	5,121,841	2,083,864

Note: Yes and No vote counts show statewide results.

Table 4: Miami-Dade Machine Randomization False Discovery Rate Tests

	Election Day			Early Voting		
item	precincts	precinct-machines	rejects	style-site-days	s-s-day-machines	rejects
Bush	734	6,976	1	1,175	7,545	1
Kerry	735	6,991	4	1,180	7,564	0
Martinez	734	6,983	0	1,205	7,690	1
Castor	736	7,001	5	1,224	7,809	2
Am. 1 yes	743	7,034	6	1,302	8,159	0
Am. 1 no	737	7,010	4	1,272	8,039	0
Am. 2 yes	742	7,031	6	1,295	8,144	4
Am. 2 no	737	7,009	1	1,228	7,901	2
Am. 3 yes	740	7,019	8	1,290	8,078	1
Am. 3 no	741	7,027	0	1,290	8,072	0
Am. 4 yes	741	7,026	2	1,313	8,209	0
Am. 4 no	739	7,017	0	1,297	8,136	0
Am. 5 yes	736	6,994	0	1,168	7,587	1
Am. 5 no	727	6,928	0	1,082	7,139	0
Am. 6 yes	742	7,031	1	1,308	8,197	0
Am. 6 no	742	7,031	0	1,271	8,061	0
Am. 7 yes	732	6,983	0	1,144	7,522	0
Am. 7 no	720	6,906	0	1,018	6,906	1
Am. 8 yes	739	7,017	0	1,272	8,043	0
Am. 8 no	735	7,000	0	1,219	7,839	2

Note: Each statistic is based on 50,000 Monte Carlo replications to compute the tail probability estimate $\hat{g}_{jl}$.

Table 5: Broward Machine Randomization False Discovery Rate Tests

	Election Day			Early Voting		
item	precincts	precinct-machines	rejects	site-days	site-day-machines	rejects
Bush	764	5,286	1	30	110	0
Kerry	765	5,289	0	30	110	0
Martinez	765	5,289	7	30	110	0
Castor	764	5,286	5	30	110	0
Am. 1 yes	767	5,293	0	30	110	0
Am. 1 no	766	5,290	0	30	110	0
Am. 2 yes	764	5,286	0	30	110	0
Am. 2 no	763	5,283	0	30	110	0
Am. 3 yes	765	5,288	2	30	110	0
Am. 3 no	765	5,288	5	30	110	0
Am. 4 yes	766	5,292	6	30	110	0
Am. 4 no	766	5,292	1	30	110	0
Am. 5 yes	757	5,266	7	30	110	1
Am. 5 no	756	5,263	1	30	110	1
Am. 6 yes	764	5,287	1	30	110	0
Am. 6 no	764	5,287	0	30	110	0
Am. 7 yes	759	5,272	4	30	110	1
Am. 7 no	757	5,266	0	30	110	6
Am. 8 yes	761	5,278	1	30	110	0
Am. 8 no	760	5,275	3	30	110	0

Note: Each statistic is based on either 10,000 or 50,000 Monte Carlo replications to compute the tail probability estimate $\hat{g}_{jl}$.

Table 6: Pasco Machine Randomization False Discovery Rate Tests

	Election Day			Early Voting		
item	precincts	precinct-machines	rejects	site-days	site-day-machines	rejects
Bush	152	1,338	2	3	35	0
Kerry	152	1,338	4	3	35	0
Martinez	152	1,338	0	3	35	0
Castor	152	1,338	2	3	35	0
Am. 1 yes	152	1,338	1	3	35	0
Am. 1 no	152	1,338	0	3	35	0
Am. 2 yes	152	1,338	0	3	35	0
Am. 2 no	152	1,338	0	3	35	0
Am. 3 yes	152	1,338	0	3	35	0
Am. 3 no	152	1,338	0	3	35	0
Am. 4 yes	152	1,338	0	3	35	0
Am. 4 no	152	1,338	0	3	35	0
Am. 5 yes	152	1,338	0	3	35	1
Am. 5 no	152	1,338	0	3	35	0
Am. 6 yes	152	1,338	0	3	35	0
Am. 6 no	152	1,338	0	3	35	0
Am. 7 yes	152	1,338	0	3	35	0
Am. 7 no	152	1,338	1	3	35	1
Am. 8 yes	152	1,338	0	3	35	0
Am. 8 no	152	1,338	0	3	35	0

Note: Each statistic is based on 10,000 Monte Carlo replications to compute the tail probability estimate $\hat{g}_{jl}$.

Table 7: Frequency of Digits according to Benford's Law

digit	0	1	2	3	4	5	6	7	8	9
first	—	.301	.176	.124	.097	.079	.067	.058	.051	.046
second	.120	.114	.109	.104	.100	.097	.093	.090	.088	.085

Table 8: Miami-Dade Election Day First-digit Benford's Law Tests

item	Benf.	equal	item	Benf.	equal
Bush	29.3	292.5	Am. 4 Yes	144.8	367.0
Kerry	39.9	287.0	Am. 4 No	119.6	605.6
Martinez	35.6	273.8	Am. 5 Yes	115.4	122.2
Castor	22.0	304.7	Am. 5 No	27.6	623.4
Am. 1 Yes	86.2	290.5	Am. 6 Yes	98.8	395.0
Am. 1 No	80.5	636.2	Am. 6 No	84.0	532.9
Am. 2 Yes	95.6	362.5	Am. 7 Yes	130.3	112.7
Am. 2 No	60.0	722.7	Am. 7 No	49.9	582.8
Am. 3 Yes	60.5	401.3	Am. 8 Yes	123.0	210.6
Am. 3 No	51.5	496.5	Am. 8 No	102.6	831.1

Note: $n = 757$ precincts. Each statistic is the Pearson chi-squared statistic, with eight degrees of freedom.

Table 9: Miami-Dade Election Day Second-digit Benford's Law Tests

item	precincts ($n = 757$) Benf.	precincts ($n = 757$) equal	machines ($n = 5,326$) Benf.	machines ($n = 5,326$) equal	precinct-machines ($n = 7,064$) Benf.	precinct-machines ($n = 7,064$) equal
Bush	7.9	10.8	28.0	20.5	17.2	39.5
Kerry	9.5	14.4	61.8	10.0	44.0	13.1
Martinez	8.9	10.8	33.4	11.9	11.5	29.2
Castor	12.0	12.8	44.5	15.6	12.7	43.5
Am. 1 Yes	2.5	8.0	72.4	10.3	43.6	12.6
Am. 1 No	5.5	15.5	73.9	9.2	19.8	31.9
Am. 2 Yes	16.7	23.6	68.5	3.5	38.7	27.3
Am. 2 No	7.2	16.4	49.5	17.3	11.9	48.8
Am. 3 Yes	3.3	8.5	98.4	9.2	78.0	5.5
Am. 3 No	12.9	12.7	76.9	9.0	25.7	26.8
Am. 4 Yes	3.3	9.0	49.1	5.8	43.5	14.4
Am. 4 No	5.7	15.4	89.5	5.4	25.4	15.3
Am. 5 Yes	17.9	19.6	81.4	3.9	57.6	2.9
Am. 5 No	5.8	23.3	5.9	56.8	25.6	135.6
Am. 6 Yes	4.3	10.2	50.3	5.8	29.7	16.3
Am. 6 No	9.1	11.3	47.3	6.5	15.3	30.8
Am. 7 Yes	17.1	16.0	51.7	21.0	53.2	21.1
Am. 7 No	8.4	16.5	78.9	220.0	136.7	318.7
Am. 8 Yes	12.7	25.3	69.6	1.5	54.2	8.3
Am. 8 No	6.5	10.6	67.8	13.9	23.2	29.1

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom.

Table 10: Second-digit Benford's Law Tests with Simulated Vote Counts

		Mean	500 precincts		750 precincts		1,000 precincts	
Size	Split	Votes	Benf.	equal	Benf.	equal	Benf.	equal
250	0.1	54.2	14.6	31.9	17.8	43.7	20.0	54.0
	0.2	62.5	13.9	30.8	17.9	43.2	19.8	52.3
	0.3	70.8	14.8	32.1	17.9	42.5	20.7	54.0
	0.4	79.2	16.0	33.0	19.6	46.0	21.5	56.1
	0.5	87.5	17.4	34.3	20.0	44.7	23.8	56.4
	0.6	95.8	13.5	24.7	14.8	29.3	17.6	36.9
500	0.1	108.3	9.4	12.4	9.8	14.9	10.0	16.4
	0.2	125.0	9.2	15.2	8.9	15.8	8.8	18.4
	0.3	141.7	10.3	13.2	10.0	13.7	10.9	17.4
	0.4	158.3	10.8	10.1	11.4	10.6	12.2	12.2
	0.5	175.0	11.1	10.5	11.0	10.7	13.1	11.8
	0.6	191.7	12.3	10.5	13.1	9.8	14.4	10.1
750	0.1	162.5	10.3	11.0	10.8	11.6	11.0	12.0
	0.2	187.5	9.6	11.3	10.2	12.1	12.4	14.2
	0.3	212.5	11.8	9.9	11.4	10.1	14.3	10.4
	0.4	237.5	12.4	9.2	12.7	9.4	15.5	9.4
	0.5	262.5	12.2	8.6	14.7	9.3	17.2	9.5
	0.6	287.5	13.1	9.3	14.2	9.1	17.0	9.3
1000	0.1	216.7	10.4	11.4	10.6	11.9	12.8	13.0
	0.2	250.0	12.3	9.8	12.9	9.6	14.7	10.7
	0.3	283.3	12.2	9.7	15.5	9.6	17.1	9.4
	0.4	316.7	13.2	8.9	15.2	9.4	16.6	8.9
	0.5	350.0	13.4	8.6	16.4	8.4	18.9	9.4
	0.6	383.3	13.5	9.5	15.3	8.5	17.5	9.1
1250	0.1	270.8	9.8	15.7	10.5	18.2	10.1	23.1
	0.2	312.5	9.1	12.3	10.5	13.9	10.9	17.1
	0.3	354.2	10.1	11.3	11.2	14.3	12.0	16.1
	0.4	395.8	11.2	13.2	12.2	15.2	13.1	16.5
	0.5	437.5	11.6	14.4	12.7	18.1	14.1	19.2
	0.6	479.2	11.6	15.9	13.5	20.8	14.9	22.0
1500	0.1	325	9.7	17.0	8.9	19.8	9.8	25.7
	0.2	375	9.1	16.1	9.9	19.1	9.9	23.0
	0.3	425	9.3	16.9	10.0	21.4	10.7	26.5
	0.4	475	11.2	22.1	11.0	25.3	12.1	31.4
	0.5	525	14.7	29.9	18.6	42.6	21.0	52.0
	0.6	575	27.2	52.7	33.9	70.3	43.7	93.2

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom, averaged over 100 Monte Carlo replications.

Table 11: Second-digit Benford's Law Tests with Simulated Vote Counts

Size	Split	Mean Votes	500 precincts Benf.	500 precincts equal	750 precincts Benf.	750 precincts equal	1,000 precincts Benf.	1,000 precincts equal
1750	0.1	379.2	9.2	18.0	9.7	23.6	10.2	28.1
	0.2	437.5	9.8	19.7	11.1	27.5	11.3	33.8
	0.3	495.8	12.8	28.1	14.8	38.1	15.6	44.1
	0.4	554.2	16.2	35.5	20.6	50.9	26.0	66.0
	0.5	612.5	27.0	54.6	35.9	77.3	41.0	94.1
	0.6	670.8	41.2	76.9	55.4	107.2	75.8	148.0
2000	0.1	433.3	10.3	21.1	11.3	28.0	12.2	34.9
	0.2	500.0	12.2	26.6	15.7	38.7	17.9	48.4
	0.3	566.7	15.0	33.8	20.5	50.2	24.3	63.9
	0.4	633.3	20.5	43.6	25.2	58.6	30.4	75.0
	0.5	700.0	26.3	53.1	34.9	74.7	45.2	99.8
	0.6	766.7	35.2	64.7	48.2	91.9	63.1	121.8
2250	0.1	487.5	14.9	31.8	17.3	43.2	23.0	60.0
	0.2	562.5	17.1	36.4	19.1	47.6	23.3	61.5
	0.3	637.5	17.9	39.1	21.4	51.5	27.2	68.8
	0.4	712.5	19.8	41.7	26.6	60.2	28.9	71.0
	0.5	787.5	23.5	47.5	31.0	67.5	42.8	93.1
	0.6	862.5	23.4	41.3	29.6	55.9	36.2	72.9
2500	0.1	541.7	17.4	37.0	20.0	48.3	25.9	64.1
	0.2	625.0	17.4	36.7	20.4	47.4	24.9	62.5
	0.3	708.3	17.2	35.2	20.5	47.4	28.2	66.3
	0.4	791.7	17.4	35.9	22.6	50.6	26.7	63.7
	0.5	875.0	18.7	36.6	23.9	50.5	28.8	64.5
	0.6	958.3	14.6	24.0	17.5	31.0	20.5	39.0
2750	0.1	595.8	14.9	30.7	18.4	41.5	21.4	50.8
	0.2	687.5	15.6	28.8	19.3	40.5	22.6	50.2
	0.3	779.2	16.3	30.2	18.3	37.4	21.2	47.3
	0.4	870.8	13.7	27.7	16.4	36.2	19.3	47.6
	0.5	962.5	12.6	21.4	15.9	30.6	19.5	38.7
	0.6	1054.2	11.0	14.8	12.3	18.7	13.9	21.1
3000	0.1	650	13.5	23.3	14.7	29.7	16.4	36.1
	0.2	750	12.2	19.7	14.6	27.5	16.2	32.4
	0.3	850	11.8	18.0	12.2	21.9	15.1	26.6
	0.4	950	10.6	19.5	11.4	23.9	11.4	29.1
	0.5	1050	12.0	17.0	11.4	18.6	11.6	21.0
	0.6	1150	11.1	12.6	11.5	13.4	11.7	16.0

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom, averaged over 100 Monte Carlo replications.

Table 12: Miami-Dade Early Voting Second-digit Benford's Law Tests

item	site-style-days ($n = 5,186$) Benf.	equal	machines ($n = 727$) Benf.	equal	site-style-day-machines ($n = 33,126$) Benf.	equal
Bush	10.1	44.9	23.5	20.9	130.3	391.4
Kerry	17.3	60.4	61.7	12.1	115.5	387.3
Martinez	14.8	48.6	32.6	17.3	107.6	357.9
Castor	9.1	42.1	43.3	18.6	93.0	336.2
Am. 1 Yes	14.1	59.9	69.6	9.8	119.7	415.4
Am. 1 No	8.7	44.1	64.8	9.8	86.3	295.7
Am. 2 Yes	17.7	65.4	58.3	2.6	83.4	334.7
Am. 2 No	20.2	71.1	41.9	16.9	92.0	292.8
Am. 3 Yes	8.2	41.4	90.8	7.6	122.7	394.8
Am. 3 No	15.3	56.7	66.1	7.8	104.8	342.1
Am. 4 Yes	7.7	40.6	47.1	11.0	87.3	338.0
Am. 4 No	14.4	60.7	83.6	5.3	108.9	351.4
Am. 5 Yes	21.9	78.3	69.2	4.6	58.4	307.5
Am. 5 No	11.0	44.8	5.7	71.6	84.4	237.8
Am. 6 Yes	12.9	56.9	55.3	11.0	105.2	368.5
Am. 6 No	9.0	37.8	44.4	9.6	126.6	374.1
Am. 7 Yes	24.6	85.0	47.8	14.9	134.2	468.3
Am. 7 No	12.0	33.9	77.4	236.4	64.5	192.7
Am. 8 Yes	13.9	61.7	68.9	2.4	96.3	377.7
Am. 8 No	6.7	28.9	63.5	15.5	79.2	261.2

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom.

Table 13: Broward Second-digit Benford's Law Tests

	Election Day				Early Voting			
	precincts ($n = 775$)		machines ($n = 5,307$)		styles ($n = 150$)		machines ($n = 190$)	
item	Benf.	equal	Benf.	equal	Benf.	equal	Benf.	equal
Bush	9.6	6.6	23.4	25.6	9.1	12.2	8.4	9.5
Kerry	21.2	12.4	79.7	6.5	21.4	24.8	10.5	17.6
Martinez	10.7	8.3	28.2	20.1	6.6	9.8	5.2	8.6
Castor	13.6	5.9	69.7	11.4	9.2	6.7	11.4	17.5
Am. 1 Yes	24.1	16.3	31.2	8.5	10.1	12.2	14.9	10.0
Am. 1 No	17.1	18.1	60.3	8.4	7.0	3.7	7.0	7.2
Am. 2 Yes	12.2	7.3	47.5	21.7	13.6	11.7	19.4	16.8
Am. 2 No	11.6	22.4	47.6	18.8	8.7	9.8	4.8	3.9
Am. 3 Yes	7.4	6.4	65.8	9.1	8.1	11.8	11.0	14.9
Am. 3 No	24.9	6.7	40.5	11.7	11.9	17.7	5.4	4.6
Am. 4 Yes	9.8	7.7	61.3	5.8	14.4	15.5	14.2	22.7
Am. 4 No	8.6	16.2	55.8	10.1	4.7	10.1	10.5	8.2
Am. 5 Yes	7.9	8.8	76.9	17.5	13.8	13.0	15.6	20.9
Am. 5 No	7.4	20.6	24.8	113.4	5.2	4.1	9.7	8.4
Am. 6 Yes	19.4	9.9	84.9	10.3	4.4	4.4	11.9	16.8
Am. 6 No	6.2	10.9	43.7	5.6	7.8	10.1	16.6	16.4
Am. 7 Yes	13.1	16.7	72.1	6.6	5.0	8.6	44.0	64.2
Am. 7 No	14.3	44.3	157.7	346.9	8.9	9.6	5.7	8.7
Am. 8 Yes	7.1	3.8	74.6	6.3	4.3	6.2	27.9	42.9
Am. 8 No	13.9	26.1	15.9	21.7	6.7	7.3	4.0	7.7

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom. In Broward, on election day each machine recorded votes for only one precinct. In the early voting data the number of votes on each style-machine combination was too small (mean = 16.7, median = 2) to support analysis for those combinations.

Table 14: Pasco Second-digit Benford's Law Tests

	Election Day				Early Voting	
	precincts ($n = 152$)		machines ($n = 1,338$)		machine-precincts ($n = 372$)	
item	Benf.	equal	Benf.	equal	Benf.	equal
Bush	6.9	5.6	16.4	16.2	14.6	23.8
Kerry	4.0	3.5	22.9	21.7	19.0	25.2
Martinez	6.5	3.7	30.6	6.4	13.4	24.3
Castor	11.2	10.5	40.5	7.7	14.7	20.7
Am. 1 Yes	9.0	10.4	24.1	11.3	5.4	10.5
Am. 1 No	7.0	5.1	9.8	5.0	18.6	28.3
Am. 2 Yes	5.4	4.8	28.6	10.3	9.6	16.2
Am. 2 No	8.6	12.7	15.8	1.9	10.4	17.7
Am. 3 Yes	10.4	9.3	34.6	11.0	12.5	18.6
Am. 3 No	8.5	4.4	10.1	16.2	13.1	19.2
Am. 4 Yes	6.0	8.4	20.7	2.8	8.6	14.7
Am. 4 No	8.6	5.2	19.8	9.3	21.5	33.4
Am. 5 Yes	3.6	9.4	16.6	8.2	11.9	20.9
Am. 5 No	3.8	6.4	10.2	19.1	10.3	17.2
Am. 6 Yes	12.8	15.5	33.5	7.7	10.5	18.7
Am. 6 No	4.4	4.7	20.1	10.0	14.4	16.4
Am. 7 Yes	29.5	43.3	20.5	18.3	14.1	22.3
Am. 7 No	5.1	7.2	19.9	10.7	5.2	6.9
Am. 8 Yes	8.0	13.8	16.5	7.7	6.3	8.6
Am. 8 No	8.0	14.6	29.9	6.6	11.1	18.1

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom. In Pasco, on election day each machine recorded votes for only one precinct. In Pasco there were only 16 early voting "precincts," too few to support analysis for those units.

Table 15: Simulated Repeaters

Size	Add	500 precincts Benf.	500 precincts equal	1,000 precincts Benf.	1,000 precincts equal
500	0.05	9.1	12.0	8.7	12.3
	0.10	8.8	13.7	9.9	19.0
	0.15	9.2	18.0	9.9	28.0
	0.20	14.5	17.8	19.5	21.6
	0.25	29.6	16.2	43.4	18.9
1000	0.05	11.4	12.6	10.7	13.3
	0.10	11.5	7.9	16.7	10.3
	0.15	15.2	11.6	18.7	12.1
	0.20	12.3	10.9	13.3	11.4
	0.25	12.5	14.8	16.5	18.3
1500	0.05	9.7	17.8	10.6	24.7
	0.10	7.8	15.7	11.4	28.2
	0.15	9.8	21.2	13.4	35.7
	0.20	18.1	39.1	25.4	66.0
	0.25	26.4	54.0	52.7	111.9
2000	0.05	12.6	26.2	23.0	57.2
	0.10	18.3	39.4	31.0	74.8
	0.15	22.0	44.1	29.5	70.9
	0.20	21.2	41.8	31.8	71.1
	0.25	20.2	35.8	33.3	68.7

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom, averaged over 25 Monte Carlo replications. Split = .1. For each size, the mean number of votes for the candidate before the repeaters are added is: 500, 108.3; 1000, 216.7; 1500, 325; 2000, 433.3.

Table 16: Simulated Repeaters with Thresholds

		500 precincts				1,000 precincts			
		Below Threshold		Above Threshold		Below Threshold		Above Threshold	
Size	Add	Benf.	equal	Benf.	equal	Benf.	equal	Benf.	equal
500	0.05	13.2	24.3	19.4	13.0	19.5	43.2	25.8	16.1
	0.10	17.5	30.5	18.9	13.4	25.9	53.5	34.4	22.3
	0.15	17.5	27.7	16.3	18.7	29.5	50.5	24.0	27.7
	0.20	14.6	15.3	8.6	12.3	18.9	20.1	9.5	17.5
	0.25	15.5	11.6	16.7	12.2	24.2	15.9	23.0	14.0
1000	0.05	12.8	15.3	13.0	11.5	18.1	22.3	17.6	16.3
	0.10	13.1	7.3	10.5	11.3	18.6	9.0	11.9	18.1
	0.15	12.3	8.4	13.8	15.1	19.6	9.6	20.0	26.5
	0.20	15.1	8.4	10.3	16.4	22.3	10.4	13.5	28.2
	0.25	15.2	10.5	15.2	21.8	19.7	12.8	21.9	36.4
1500	0.05	9.3	11.4	11.9	26.8	10.5	14.1	13.5	38.9
	0.10	11.0	12.6	11.0	25.6	10.7	13.2	16.3	44.0
	0.15	7.6	11.4	13.8	31.1	11.3	17.4	20.0	54.6
	0.20	9.4	13.3	22.4	47.7	8.4	16.6	42.6	96.2
	0.25	10.3	12.3	41.8	77.8	10.3	15.7	72.6	142.9
2000	0.05	9.5	19.0	15.8	34.7	10.8	26.6	24.3	60.8
	0.10	8.1	14.6	21.5	46.0	10.9	29.1	29.6	74.4
	0.15	8.8	17.6	24.5	48.2	11.7	29.4	38.4	87.1
	0.20	7.9	14.8	21.9	42.9	9.5	26.4	42.3	88.0
	0.25	11.4	20.9	23.1	42.4	10.3	25.5	38.7	75.9

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom, averaged over 25 Monte Carlo replications. Split = .1. For each size, the mean number of votes for the candidate before the repeaters are added is: 500, 108.3; 1000, 216.7; 1500, 325; 2000, 433.3.

Table 17: Simulated Proportional Adjustments

		500 precincts		1,000 precincts	
Size	Prop.	Benf.	equal	Benf.	equal
500	0.8	12.9	18.0	18.0	29.5
	0.85	10.7	14.4	9.3	18.2
	0.9	7.7	13.0	7.7	16.0
	0.95	9.4	11.4	8.6	13.0
	1.05	10.4	11.1	10.1	14.8
	1.1	9.6	14.8	10.2	15.4
	1.15	9.7	10.3	13.7	13.2
	1.2	11.3	14.4	13.2	16.5
1000	0.8	16.2	18.0	15.6	19.6
	0.85	10.3	10.0	12.7	11.8
	0.9	10.9	10.7	11.3	10.1
	0.95	10.7	12.1	11.4	11.9
	1.05	9.9	10.6	10.9	11.7
	1.1	10.3	14.2	10.4	19.2
	1.15	11.0	14.6	10.6	15.2
	1.2	9.9	15.2	10.0	19.4
1500	0.8	10.5	15.0	13.7	27.7
	0.85	10.0	15.7	10.0	23.2
	0.9	9.5	17.0	10.0	24.1
	0.95	10.1	17.7	10.1	24.9
	1.05	9.2	16.9	8.1	23.1
	1.1	9.6	18.4	9.5	27.0
	1.15	10.6	19.5	9.1	25.6
	1.2	10.1	20.6	10.2	28.8
2000	0.8	10.5	20.5	11.1	29.7
	0.85	8.6	16.3	9.6	27.4
	0.9	9.5	20.0	12.3	31.2
	0.95	8.4	17.8	10.3	30.4
	1.05	12.9	26.8	16.0	45.2
	1.1	15.5	33.1	23.2	59.8
	1.15	16.8	34.3	27.3	69.2
	1.2	18.3	39.5	23.5	61.8

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom, averaged over 25 Monte Carlo replications. Split = .1. For each size, the mean number of votes for the candidate before the repeaters are added is: 500, 108.3; 1000, 216.7; 1500, 325; 2000, 433.3.

Table 18: Simulated Proportional Adjustments with Thresholds

		500 precincts				1,000 precincts			
		Below Threshold		Above Threshold		Below Threshold		Above Threshold	
Size	Prop.	Benf.	equal	Benf.	equal	Benf.	equal	Benf.	equal
500	0.8	17.9	18.3	10.1	14.4	23.6	24.4	12.3	22.3
	0.85	10.7	10.4	7.1	11.5	11.3	10.6	8.4	16.8
	0.9	10.9	9.8	9.5	15.1	12.0	10.3	9.5	21.0
	0.95	12.5	10.2	11.3	15.4	17.3	13.5	10.5	19.0
	1.05	15.3	20.8	14.2	12.5	21.4	30.8	22.5	18.1
	1.1	22.3	33.0	15.5	11.6	33.7	53.8	26.0	18.3
	1.15	17.9	26.2	17.8	12.2	22.1	36.3	23.5	13.4
	1.2	21.5	31.1	16.6	10.7	33.6	52.0	26.5	15.4
1000	0.8	17.2	23.3	14.3	10.1	23.3	34.2	22.6	13.8
	0.85	11.9	14.1	11.5	9.5	16.0	22.2	17.9	11.9
	0.9	10.4	10.7	9.4	10.5	10.8	12.1	11.8	10.0
	0.95	9.1	9.7	10.3	9.8	10.9	12.0	13.0	11.3
	1.05	11.2	17.4	14.0	8.3	16.5	28.5	21.4	11.2
	1.1	14.5	23.1	11.9	8.4	19.1	35.9	20.7	12.4
	1.15	12.1	17.8	13.4	10.6	14.7	25.2	17.5	12.6
	1.2	14.0	21.8	11.7	10.9	17.2	29.3	17.4	14.4
1500	0.8	13.9	20.8	12.4	12.7	17.9	30.7	19.4	17.2
	0.85	11.3	15.6	11.9	14.1	12.9	20.5	14.9	20.7
	0.9	10.3	12.5	11.6	17.3	10.2	12.5	11.5	21.4
	0.95	10.0	10.9	8.7	13.2	11.4	15.2	11.6	22.2
	1.05	8.9	11.8	9.6	14.3	10.2	17.7	9.0	19.3
	1.1	9.8	13.2	8.3	14.3	11.6	18.2	11.9	24.6
	1.15	10.2	10.3	10.9	18.8	13.0	13.9	12.5	30.7
	1.2	10.8	11.2	12.8	22.9	14.3	12.1	15.4	36.6
2000	0.8	11.0	13.8	11.2	17.0	14.9	21.2	10.7	22.1
	0.85	11.3	13.2	10.4	19.1	11.7	14.5	12.4	25.4
	0.9	11.0	10.1	13.6	24.9	12.6	10.8	15.4	34.9
	0.95	12.2	9.8	10.7	20.3	11.4	10.6	13.4	32.7
	1.05	11.1	11.4	16.5	30.3	9.8	11.9	25.0	51.9
	1.1	10.3	11.1	21.4	39.1	9.5	11.2	34.6	68.6
	1.15	12.3	10.2	25.3	46.4	15.3	10.5	50.0	95.2
	1.2	12.4	12.8	23.2	42.7	8.7	9.2	38.2	76.9

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom, averaged over 25 Monte Carlo replications. Split = .1. For each size, the mean number of votes for the candidate before the repeaters are added is: 500, 108.3; 1000, 216.7; 1500, 325; 2000, 433.3.

Table 19: Simulated Counts for Miami-Dade Precincts and Machines

	precincts		machines	
Split	Benf.	equal	Benf.	equal
0.1	9.5	14.5	9.5	69.5
0.2	9.4	14.3	10.3	61.0
0.3	9.6	15.9	12.6	45.2
0.4	9.1	13.5	16.8	35.1
0.5	8.8	12.6	21.8	26.6
0.6	11.1	12.3	29.2	25.0
0.7	9.4	13.1	33.0	18.9

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom, averaged over 25 Monte Carlo replications.

Table 20: Calibrated Simulated Counts for Miami-Dade Precincts and Machines

	Bush				Kerry			
Calibrated	precincts		machines		precincts		machines	
Parameters	Benf.	equal	Benf.	equal	Benf.	equal	Benf.	equal
actual precincts	7.9	10.8	16.3	35.7	9.5	14.4	36.7	19.1
splits	10.4	18.2	19.4	109.3	9.2	18.6	16.0	103.0
splits and betas	9.8	15.2	11.1	48.6	9.4	14.8	12.2	49.1

Note: Each statistic is the Pearson chi-squared statistic, with nine degrees of freedom. The simulated statistics are averaged over 100 Monte Carlo replications.

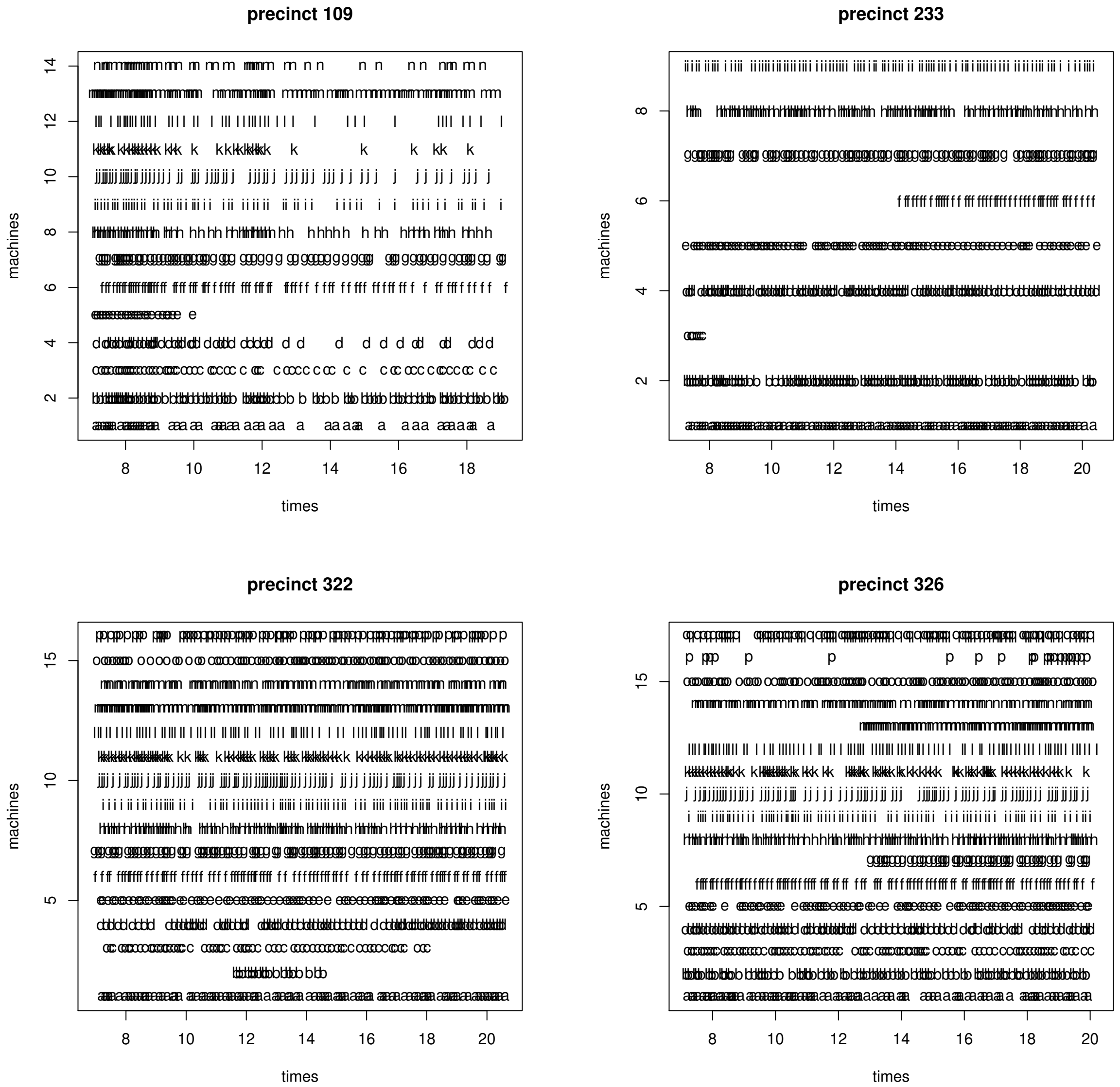

Figure 1: Times (Resolved to the Second and Shown on a 24-Hour Clock) When Votes Were Cast on Machines in Selected Precincts on Election Day, Miami-Dade County

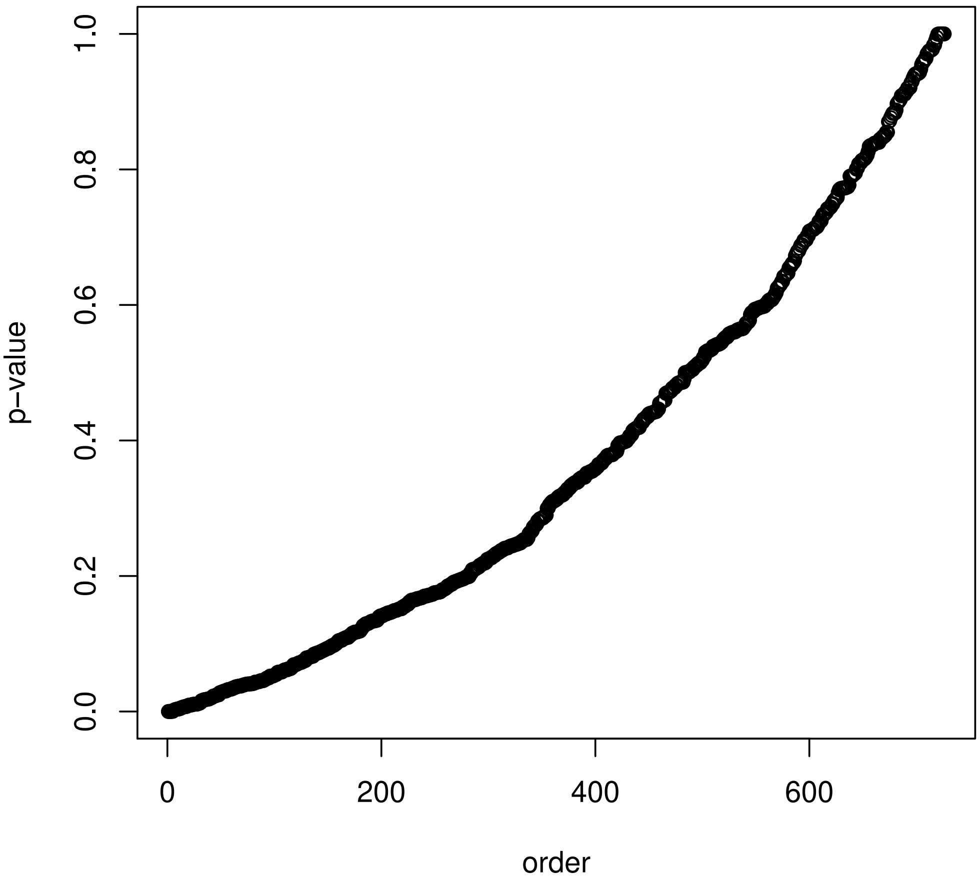

Figure 2: Miami-Dade Election Day Voting Machine Randomization Test Tail Probabilities

Actual

Calibrated Simulation

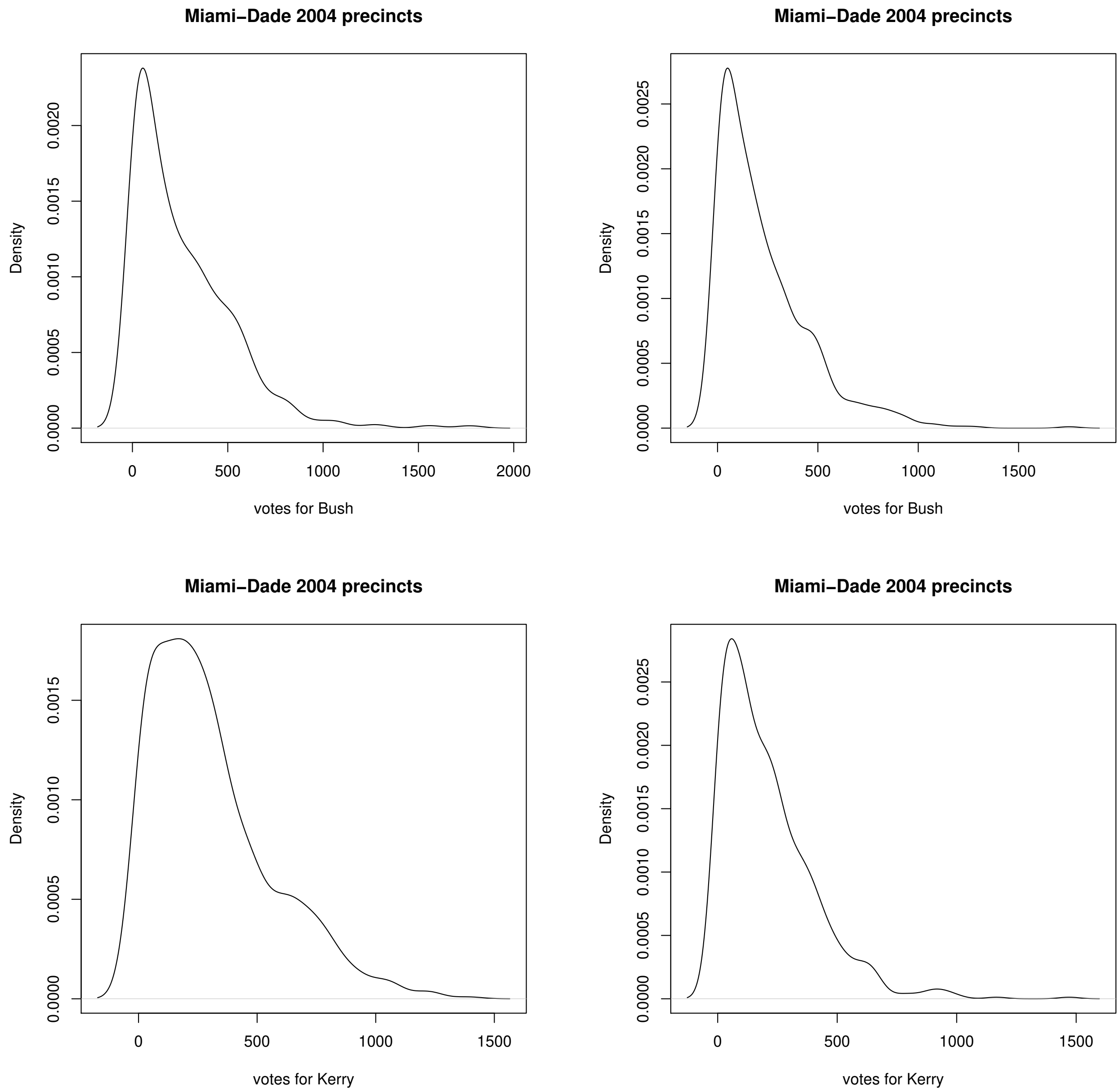

Figure 3: Miami-Dade Election Day Precinct Vote Count Distributions