Election Forensics: The Second-digit Benford's Law Test and Recent American Presidential Elections *

Walter R. Mebane, Jr.†

November 3, 2006

*An earlier version of this paper was presented at the Election Fraud Conference, Salt Lake City, Utah, September 29–30, 2006. I thank Herbie Ziskend, Charlie Gibbons and Gideon Weissman for assistance.

†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).

Abstract

Election Forensics: The Second-digit Benford's Law Test and Recent American Presidential Elections

While the technology to conduct elections continues to be imperfect, it is useful to investigate methods for detecting problems that may occur. A method that seems to have many good properties is to test whether the second digits of reported vote counts occur with the frequencies specified by Benford's Law. I illustrate use of this test by applying it to precinct-level votes reported in recent American presidential elections. The test is significant for votes reported from some notorious places. But the test is not sensitive to distortions we know significantly affected many votes. In particular, the test does not indicate problems for Florida in 2000. Regarding Ohio in 2004, the test does not overturn previous judgments that manipulation of reported vote totals did not determine the election outcome, but it does suggest there were significant problems in the state. The test is worth taking seriously as a statistical test for election fraud.

Arguably we are not much closer than we were one hundred years ago to understanding how to administer elections that not only are secure and fair but are widely believed to be secure and fair. As long as there have been elections there have been election scandals, and certainly throughout the history of the United States (Gumbel 2005). Notoriously, serious defects in election administration produced the wrong outcome in the 2000 American election for president (Wand, Shotts, Sekhon, Mebane, Herron, and Brady 2001; Mebane 2004b). Responses to the 2000 election controversy have in some ways created as many problems as they have solved. In particular, the events of 2000 sparked a rush to replace older mechanical voting technologies with machines based on electronic computers. Some states made such changes on their own, notably Florida (MacManus 2004), while others were prompted to change by provisions of the Help America Vote Act of 2002 that made use of punchcard ballots and lever machines illegal and provided funds to help pay for their replacement.

Debates about the accuracy and security of different voting technologies began in the nineteenth century and continue to the present (Saltman 2006). In 1934, Joseph Harris wrote about a primary defect of paper ballots: "The counting of paper ballots, often lasting far into the night, and made by tired and frequently incompetent persons, is highly conducive to mistakes and frauds. Many election officers and men in public life have realized the inherent defects of this procedure and have sought to remedy it" (Harris 1934, 261). But in 2006, Aviel Rubin lamented the insecurity of a wholly electronic system being used in Maryland: "All of the votes from our entire precinct were right there in the palm of my hand. I could have substituted those five [memory] cards with five identical but bogus cards from my pocket, changing all the ballots, because Diebold did not protect the data with appropriate cryptographic measures" (Rubin 2006, 256). It seems unlikely that technological developments alone will solve the problems and resolve the questions many have about election administration, at least not in the foreseeable future.

While the technology to conduct elections continues to be imperfect, it is useful to investigate methods for detecting problems that may occur. The class of such methods I refer to as *election forensics* are based on statistical tools and are intended to examine elections after the fact. Election forensics focus on the recorded votes, asking whether there are significant anomalies. Do the votes relate to covariates in ways we should expect, or are some votes outliers (Mebane and Sekhon 2004)? Are there other regularities the votes should exhibit? The analysis by Wand et al.

(2001) of the consequences of the butterfly ballot in the 2000 presidential election features both of these kinds of analysis. That study finds that the vote for Pat Buchanan in Palm Beach County was a significant outlier, that the vote for Buchanan on election day ballots in Palm Beach County did not relate to the vote on absentee ballots in the same way as it did in other Florida counties, and that the vote for Buchanan did not track the vote for other Reform Party candidates running in Palm Beach County.

Of course the most challenging ambition for election forensics is to be able to detect election fraud. An examination merely of recorded votes and their correlates can never by itself prove that regularities or irregularities the recorded votes may exhibit are the result of fraudulent intentions. But allegations of fraud may identify specific methods purportedly used to perpetrate the fraud, and the forensic analysis may be able to check for traces of those methods. Such an analysis may help reduce suspicions that election results are fraudulent. A study of votes cast in Ohio in the presidential election of 2004 commissioned by the Democratic National Committee documents many problems with the way the election was administered, but it does not find evidence to support charges that George W. Bush won only because tens of thousands of votes that were cast in favor of John Kerry were instead counted as votes for Bush (Mebane and Herron 2005).

The ideal method for election forensics would be one that depends neither on special assumptions about the particular political configurations contesting the election nor on any particular theory about how the election was conducted. Ruled out, for instance, would be ideas about the coalitions supporting a particular party or candidate. In general we should expect a method that is based on particular theories to be more powerful than a method that eschews such foundations, at least if the theories are correct. But any particular theory is likely also to be controversial. A diagnosis of election fraud—or of its absence—that depends on such theorizing may be only as convincing as is the theory it depends on.

An ideal method for election forensics would also be one that could be applied routinely, perhaps even automatically, without requiring special expertise or sophisticated technical judgment. Such a method might be a foundation for routine election audits. For instance, election officials might apply a simple test to publicly available information and then perform some kind of intensive manual inspection of places or equipment that performed poorly on the test. For example, all precincts might have a positive probability of undergoing an audit immediately after the election that includes verification of ballot and machine chains of custody and a full manual recount, with the selection probabilities being substantially higher where the routine test had a significant result.

While such an ideal method may well not exist, in this paper I want to illustrate the use of one possible candidate. A method that may come close to satisfying our ideal set of requirements is to test whether the second digits of reported vote counts occur with the frequencies specified by Benford's Law (Raimi 1976; Hill 1995). In Mebane (2006) I study this second-digit Benford's Law (2BL) test for vote counts. I identify a pair of flexible mechanisms that may generally characterize vote counts and that satisfy the 2BL distribution in a wide range of circumstances. I show that the 2BL test is sensitive to many patterns of vote count manipulation, including patterns that would occur in some kinds of election fraud. I argue that while the 2BL test may be generally suitable for precinct-level data, it is not useful for vote counts at the level of individual voting machines.

The 2BL test is not precisely theory free, and its suitability for a wide variety of electoral contexts has yet to be demonstrated. But it does fulfill the goal to free tests for election fraud from being bound to a particular idea about the substance of the campaigns or about the grounds for voters' decisions. The 2BL test uses only the vote counts themselves. No covariates are involved, and no statistical models need to be estimated. Given precinct-level vote count data, the test is very quick to compute (the hard part is obtaining the precinct data). The test results are not sharply diagnostic: Mebane (2006) shows the test can be triggered when votes are not being manipulated at all, and even if manipulation is occurring the test cannot indicate whether the manipulation is due to fraudulent actions.

The relationship between the 2BL test and manual recounts is unclear. While the 2BL test is far from perfect, there are also limits on the kinds of fraud a manual recount may detect. Harris (1934) discusses many kinds of fraud, but there is a basic distinction between two broad classes. One class of frauds involve miscounting the ballots. For example, Harris writes, "The old form of voting fraud—that of repeating—has largely disappeared. It is safer and cheaper to have the election officers steal the election. This may be done by turning in an election return which is not based upon an actual count of the ballots, and does not at all correspond to the votes cast" (Harris 1934, 262). The other class of frauds involves falsifying ballots: "Another method of stealing an election is to stuff the ballot box with marked ballots, writing in the poll books the names of voters who failed to vote or who have died or moved away" (Harris 1934, 262). A routine recount may uncover a fraud of the first kind, but it would do nothing to reveal a fraud of the second kind. But the 2BL test may be sensitive to either kind of fraud. A statistical test, such as the 2BL test, and a program of manual recounts may reinforce one another but they are not redundant.

This potential capacity for the 2BL test to signal frauds that a recount cannot catch is of course one of the strongest arguments in its favor during a time, such as now, when many jurisdictions are using electronic voting machines that do not produce a reliable audit trail, so that useful recounts are impossible. The Diebold system in Maryland that Rubin (2006, 256) writes about is one example. There is very little reason to believe such systems are secure. Rubin's worry about memory cards being swapped is not the most serious potential problem. If malicious software is installed on the machines, as demonstrated by Feldman, Halderman, and Felten (2006), then all the vote counts and every available electronic record may be falsified. Such falsification may be done in ways that would escape detection by the 2BL test. Neither the 2BL test nor any other statistical test is a panacea.

I illustrate use of the 2BL test by applying it to some of the precinct-level votes reported in recent American presidential elections. Because the controversies attending some of these election outcomes have been examined using other tools, this kind of survey will one hopes help build intuition about what the 2BL test can and cannot do. After briefly describing how to perform the 2BL test and considering some motivation for it, I return to Florida, 2000, to see whether the test flags any of the problems that are amply well documented to have happened there. Notwithstanding Florida's comprehensive reform of election administration after 2000, problems occurred in some places—e.g., in Miami-Dade and Broward counties during the 2002 gubernatorial election (Canedy 2002, 2003)—and allegations arose regarding suspected manipulation of the presidential votes in 2004. So I look at data from the 2004 election in Florida. Next I consider whether 2BL test results support the conclusions reached by Mebane and Herron (2005) about the 2004 election in Ohio. After that I take a look at 2BL test results for presidential votes from across the U.S. in 2000 and 2004.

The Second-digit Benford's Law Test for Vote Counts

The 2BL test for vote counts in J precincts uses the distribution of second digits shown in Table 1. Let $q_{B_2 i}$ denote the expected relative frequency with which the second digit is i. These $q_{B_2 i}$ values are the values shown for each digit in Table 1. Let d_{2i} be the number of times the second digit is i among the J precincts being considered, and let $d_2 = \sum_{i=0}^{9} d_{2i}$ denote the total number of second digits. If some precincts have vote counts less than 10, so those small counts lack a second digit, then $d_2 < J$. The statistic I use for a 2BL test is the Pearson chi-squared statistic:

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

These statistics may be compared to the chi-squared distribution with 9 degrees of freedom (χ_9^2), which has a critical value of 16.9 for a .05-level test.

*** Table 1 about here ***

Why should Benford's Law apply to the second digits of vote counts? This is a challenging question that I address more fully in Mebane (2006). Vote counts are complicated, being produced by several intersecting processes. Individuals decide whether to vote at all and, if so, they decide which candidates or ballot initiatives to choose. Laws and administrative decisions present a menu of alternative methods each person may use to vote: election day voting in person, early voting, provisional ballots or mail-in ballots; on paper, with machine assistance or using some combination. Various rules and practices constrain which methods each person may choose and how the votes recorded in each category are aggregated. Conditions vary across polling places: relatively poor administration in some locations may cause errors to occur more frequently in those places. To motivate the 2BL test we need an argument that shows that such a complex of processes in general implies that the resulting vote counts have digits that follow the 2BL distribution.

In Mebane (2006) I suggest that one way to construct such an argument is to focus on the moment just before each vote gets recorded. When all is said and done, most voters will look at each option on the ballot and have firm intentions either to select that option or not to select that option. Then for whatever reason—momentary confusion, bad eyesight, defective voting technology—a small proportion of those intended votes will not be cast or recorded correctly. A small proportion will be "mistakes." Differences in partisanship, economic class, mobilization campaigns, adminstrative rules or whatnot cause the proportion of voters who intend to choose each candidate or ballot question to vary across precincts. If such variations are combined with small probabilities of "mistakes" in making or recording each choice, then the resulting precinct vote counts will often follow the 2BL distribution. In Mebane (2006) I present simulations to illustrate that such a mixture of processes—where the support for each alternative on the ballot varies over precincts and there are "mistakes"—can produce 2BL distributions of vote count digits even when the number of voters is constant across precincts. I also show that a related mixture produces 2BL-distributed vote counts when the number of voters varies extensively across precincts. Presumably a wide variety of combinations of these kinds of mechanisms produce 2BL-distributed vote counts. Such a variety presents a rich family of processes that may correspond to what happens in real elections.

In what sense can the 2BL test detect election fraud? In Mebane (2006) I show that if we start with vote counts that follow the 2BL distribution, then various ways of adding or subtracting votes from those counts will produce a significantly large value of $X^2_{B_2}$. In particular, I present simulations to show that if votes are systematically added or subtracted in precincts where a candidate is already receiving more than the number of votes expected according to the vote-generating process, then the test statistic will tend to be large. In other words, the test detects when votes are added or subtracted in precincts where a candidate is already strong. In a scenario where the baseline, uncorrupted election process is expected to produce a tied outcome, increasing the vote counts in the manipulated districts by as little as two or three percent can trigger a significant test result. The test can also be triggered when votes are added or subtracted in precincts where, before the manipulation, the candidate is receiving less than the number of votes expected according to the vote-generating process. If the amount of manipulation is sufficiently small, the 2BL test will not signal that manipulation has occurred. The 2BL test is not sensitive to manipulations that involve adding or subtracting votes from a moderate number of precincts selected entirely at random. But the test would be triggered if somehow all the votes were replaced with counts generated using some simple random process (e.g., Poisson, binomial or negative binomial counts).

In general we will not be examining only one set of precincts or one set of vote counts. We may be interested in the sets of precincts in different counties or different electoral districts. We may want to look at the votes cast for different candidates, for different offices or for different ballot items. To get a simple omnibus test result, one could pool all the different vote counts together. But especially in the case where the test rejects the hypothesis that the second digits of all the vote counts follow the 2BL distribution, it will be more perspicuous to test each natural subset of precincts separately. Doing so may allow one to identify for which set of precincts the test is signaling a problem. So the votes recorded for a presidential candidate in all the precincts in a county may be considered a set and tested together, but each county is treated separately.

When computing the 2BL test for multiple sets of precincts, we need to adjust any assessment of statistical significance for the fact that we are looking at multiple tests. The method I use to do this is to adjust the test level applied to hypothesis tests to control the false discovery rate (FDR) (Benjamini and Hochberg 1995). Let $t = 1, \ldots, T$ index the T independent sets of precincts being tested. For instance, if we were testing the precincts in a state separately for each county, T might denote the number of counties in the state. Let the significance probability of the test statistic for each set be denoted S_t. In our case this probability is the upper tail probability of the χ^2_9 distribution. Sort the values S_t from all T sets from smallest to largest. Let $S_{(t)}$ denote these ordered values, with $S_{(1)}$ being the smallest. For a chosen test level α (e.g., $\alpha = .05$), let d be the smallest value such that $S_{(d+1)} > (d+1)\alpha/T$. This number d is the number of tests rejected by the FDR criterion. If the second digits of the vote counts in all of the sets do follow the 2BL distribution, then we should observe $d = 0$.

Florida 2000 and 2004

For the votes recorded for president in 2000, I have data for precincts in all of Florida's 67 counties.[1] These data exclude absentee ballots. Five counties have too few precincts to support a useful analysis (e.g., Baker County has eight precincts). I use the 62 counties that have at least ten precincts.

I compute the 2BL test for the votes recorded for George W. Bush and for Al Gore. I treat each county's precincts as a separate set, and I also treat separately the Bush and Gore vote

[1] Data come from Florida Legislative Staff (2001), file precinct00_pl.zip.

totals. For 62 counties and two candidates we have $T = 124$ separate test statistics. Controlling the FDR gives 30.2 as the critical value the 2BL statistic must exceed to signal a significant departure for the 2BL distribution.

None of the 2BL test statistics comes close to exceeding that FDR-controlled critical value. The largest statistic is $X^2_{B_2} = 28.7$ for the vote for Gore in Charlotte County. Nine other statistics are larger than the single-test critical value of 16.9—four statistics are for votes recorded for Bush and five are for votes recorded for Gore. None of those counties (Calhoun, Charlotte, Dixie, Hendry, Hillsborough, Levy, Manatee, Orange and St. Johns) is among those associated with the biggest controversies in 2000. Notably, Duval County has small statistic values ($X^2_{B_2} < 8$), and the statistics for Palm Beach County are not large. For Palm Beach, $X^2_{B_2}$ is 11.3 for Gore and 16.8 for Bush. In both counties high proportions of ballots were spoiled due to overvotes.[2]

For the major party presidential votes recorded in Florida in 2000, then, the 2BL test does not signal any significant problems. Clearly the test is not responding to some major distortions that happened in some of the counties. Neither the overvotes nor the undervotes that plagued voters in the state cause the test to trigger.

Let's fast forward, then, to 2004.

By 2004, all of Florida's counties used either precinct-tabulated optical scan or electronic touch-screen voting machine technology. While these and other changes significantly improved election administration and reduced the frequency of errors (MacManus 2004), allegations nonetheless arose that vote totals had been manipulated using both modalities. Allegedly the scanners that tabulated the paper ballots were hacked, so that suspiciously many registered Democrats were recorded as voting for Bush (Roig-Franzia and Keating 2004). These allegations largely evaporated in light of the finding that registered Democrats had long been voting for the Republican presidential nominee in the referent parts of Florida (Mebane 2004a). Moreover, careful comparisons between parts of the state that used different kinds of voting technology but were otherwise similar fail to turn up significant differences in voting patterns (Sekhon 2004; Wand 2004). Finally, a manual reinspection of the ballots in three of the supposedly affected

[2]Effects of the butterfly ballot on the vote recorded for Buchanan in Palm Beach are apparent: $X^2_{B_2} = 19.4$. Because Buchanan received so few votes, this result is a bit suspect. The votes for Buchanan exceed 10, so that the vote count has a second digit, in only 113 of the 534 Palm Beach County election day precincts for which the FREDS 2000 dataset (Florida Legislative Staff 2001) includes data.

counties finds no signs of manipulation (Laughlin and Kidwell 2004). At the other end there were allegations that some counties that used electronic voting machines recorded a surprisingly large number of votes for Bush (Zetter 2004). The statistical analysis supporting these allegations was widely discredited as unsound, but nonetheless the suspicions they abetted remained in the air (e.g. Miller 2005).

For computing the 2BL test in Florida in 2004, I have usable precinct data from 50 counties.[3] I compute the 2BL test for the votes recorded for Bush and for Kerry. I include the totals reported for absentee ballots and for early voting, treating these totals as if they are from separate precincts as given in the reported data. For 50 counties and two candidates we have $T = 100$ separate test statistics, which implies an FDR-controlled critical value for the 2BL statistic of 29.7.

Once again, none of the 2BL test statistics is larger than the FDR-controlled critical value. One value comes close. For the vote for Bush in Manatee County, $X^2_{B_2} = 28.5$. The next largest value is $X^2_{B_2} = 21.4$, for the vote for Bush in Collier County. In all there are eight statistics larger than the single-test critical value of 16.9—three statistics are for votes recorded for Bush and five are for votes recorded for Kerry. Because we are looking at so many different tests, however, these single-test results are not a compelling indication of departures from the 2BL distribution. For the major party presidential votes recorded at the precinct level in Florida in 2004, the 2BL test does not signal any significant problems.

Ohio 2004

When measured in terms of controversies and challenges, clearly the most important state in the 2004 American presidential election was Ohio. The state's electoral votes were pivotal in determining the Electoral College winner, and indeed the votes from the state were challenged in Congress when the electoral votes were counted (Stolberg and Dao 2005). That challenge was prompted in part by a report that documented extensive and serious difficulties voters in the state experienced due to partisan and poor election administration (House Judiciary Committee

[3] The counties are Alachua, Bay, Bradford, Brevard, Broward, Calhoun, Charlotte, Citrus, Clay, Collier, Columbia, Dixie, Duval, Escambia, Flagler, Gadsden, Gilchrist, Gulf, Hamilton, Hardee, Hendry, Highlands, Hillsborough, Holmes, Indian River, Jackson, Lake, Lee, Leon, Levy, Manatee, Marion, Martin, Miami-Dade, Nassau, Okaloosa, Okeechobee, Orange, Palm Beach, Pasco, Pinellas, Putnam, Sarasota, Seminole, St. Johns, Sumter, Suwannee, Taylor, Wakulla and Walton counties. Data are from Dave Leip (http://www.uselectionatlas.org).

Democratic Staff 2005). The Democratic National Committee (DNC) sponsored a study to further document and diagnose what happened in Ohio (Voting Rights Institute 2005).

One of the principal findings of the DNC study is that an examination of precinct vote totals from across the state produces "strong evidence against the claim that widespread fraud systematically misallocated votes from Kerry to Bush" (Mebane and Herron 2005, 2). Specifically this claim refers to the results of matching precincts and wards that did not change boundaries between 2002 and 2004, and then robustly estimating an "overdispersed binomial regression model that has the proportion voting for Kerry depending on the proportion voting for the Democratic candidate for governor (Tim Hagan) in the 2002 election" (Mebane and Herron 2005, 13–14). Using D2002 to represent the proportion voting for Hagan (versus Republican candidate Bob Taft) and $\text{logit}(p) = \log(p/(1-p))$ to denote the log-odds function, the model uses the following linear predictor:

$$Z_i = d_0 + d_1 \text{logit}(\text{D2002}_i) \ . \tag{1}$$

Mebane and Herron state, "If the vote for Kerry were the same as the vote for Hagan except uniformly higher, then we would have $d_0 > 0$ and $d_1 = 1$" (2005, 14). The stated parameter values are almost but not quite what they find. In Table 2 I reproduce the parameter estimates they report (Mebane and Herron 2005, 77, Table 30). The estimates for d_0 are positive, and "the estimate for d_1 is not substantially different from 1.0 in either the precinct analysis or the ward analysis" (Mebane and Herron 2005, 14). But careful examination shows the difference between the estimate for d_1 and the value 1.0 to be statistically significant. Does this small but significant difference point to a big hole in Mebane and Herron's substantive conclusion?

*** Table 2 about here ***

Mebane and Herron's explanation for equation (1) is a bit terse and does not fully articulate the rationale for the parameter values they focus on. Before I consider what the 2BL test may suggest about the matter, let me try to explain their model more explicitly.

First let's think about the votes for Hagan and for Taft in the 2002 gubernatorial election. Suppose for simplicity that the differences in support for Hagan and Taft across precincts can be largely explained in terms of a single variable x—call this "precinct net party strength"—so that the number of votes expected for each candidate in precinct i is well described by the model

$$E(\text{number for Hagan}_i) = n_i \exp(a + bx_i)$$
$$E(\text{number for Taft}_i) = n_i \exp(c + dx_i)\,.$$

Here n_i is a measure of the number of potential voters in precinct i, a, b, c and d are constants with $\text{sign}(b) = -\text{sign}(d)$, and x_i varies from precinct to precinct. The variables n and x do not perfectly capture the support for each candidate, so the actual number of votes each receives differs somewhat from the expected values. But given the value of x, the proportion of votes expected to go to Hagan is

$$E(\text{D2002}_i) = \frac{\exp(a + bx_i)}{\exp(a + bx_i) + \exp(c + dx_i)} = \frac{1}{1 + \exp[-(A + Bx_i)]}\,,$$

where $A = a - c$ represents the difference between the candidates' overall base levels of support and $B = b - d$ represents the net degree to which their support varies across precincts in relation to x. Applying the logit function to this expectation recovers the linear predictor we would use in a binomial model if we could observe x, namely $\text{logit}[E(\text{D2002}_i)] = A + Bx_i$.

If the vote for Kerry were the same as the vote for Hagan and the vote for Bush were the same as the vote for Taft, then the same model would apply to Kerry's share of the votes as applies to Hagan's. In that case we might hope that our imagined model for $E(\text{D2002}_i)$ does not depart too far from the observable values D2002_i, so that $\text{logit}(\text{D2002}_i) \approx A + Bx_i$ is a good approximation. Since we are not committing to any particular definition for the unobservable variable x, the idea that the approximation is a good one should be easy to accept. But then it follows that if Kerry's vote share were the same as Hagan's, then the linear predictor in the model for Kerry's vote share should be the same as in the model for Hagan's vote share. Hence in equation (1) we would have $d_0 = 0$ and $d_1 = 1$. That is, using D2004 to represent the proportion voting for Kerry instead of Bush, we would have

$$E(\text{D2004}_i) = \frac{1}{1 + \exp[-(A + Bx_i)]} \approx \frac{1}{1 + \exp[-\text{logit}(\text{D2002}_i)]}\,.$$

In saying that Kerry's support may be "uniformly higher" than Hagan's, the idea is that the difference between the overall base levels of support for Kerry and Bush may be more favorable to Kerry than the corresponding difference between Hagan and Taft is to Hagan. This may happen even while Kerry's and Bush's support varies across precincts in relation to x in the same way as does Hagan's and Taft's. That is, if we use G to denote the difference between Kerry's and Bush's base levels of support and H to denote the net degree to which their support varies in relation to x, so that

$$E(\text{D2004}_i) = \frac{1}{1 + \exp[-(G + Hx_i)]} ,$$

then it may be that $G > A$ while $H = B$. In this case we would have

$$E(\text{D2004}_i) \approx \frac{1}{1 + \exp\{-[G - A + \text{logit}(\text{D2002}_i)]\}} ,$$

which is to say, in equation (1), $d_0 = G - A > 0$ and $d_1 = 1$.

Of course, both Kerry and Bush received more votes than, respectively, Hagan and Taft, which is to say that overall voter turnout was higher in 2004 than in 2002. The increase in turnout reflects not only the difference that generally occurs between midterm and presidential election years, but also the intensive mobilization efforts undertaken in Ohio in 2004 by the candidates, by the political parties and by other groups. It is possible, but of course not necessary, that the mobilization worked in such a way that each candidate was able to increase turnout more in precincts where his party's base support was already stronger. In terms of a simple model expressing the support for Kerry or Bush in terms of the notional 2002 precinct net party strength variable x, the number of votes for each candidate in each precinct might be related to x_i through coefficients u and v, with $\text{sign}(u) = -\text{sign}(v)$ and $H = u - v$. Intense mobilization that was more effective in precincts where a party was already strong would mean that $|u| > |b|$ and $|v| > |d|$, in which case $H/B > 1$.

In this case, in the linear predictor of equation (1) we would no longer expect $d_1 = 1$. That is,

$$\begin{aligned} E(\text{D2004}_i) &= \frac{1}{1 + \exp\{-[G - (H/B)A + (H/B)(A + Bx_i)]\}} \\ &\approx \frac{1}{1 + \exp\{-[G - (H/B)A + (H/B)\text{logit}(\text{D2002}_i)]\}} . \end{aligned}$$

So in general in equation (1) we have $d_1 = H/B$ and $d_0 = G - (H/B)A$. If the parties in 2004 tended to mobilize more effectively in precincts where they were already strong in 2002, then we should see $d_1 > 1$.

The force of this more explicit motivation for equation (1), then, is to support a claim that the estimated values for d_1 that are significantly larger than 1.0 reflect the tremendous voter mobilization efforts undertaken on behalf of the candidates. If Bush tended to recruit new voters more effectively in precincts where he was already strong, or if Kerry tended to add voters more in places where he was already strong, then $d_1 > 1$ is what we should expect.

Evidently the analysis that refers to equation (1) depends on an elaborate skein of modelling, even though the ideas it expresses are fairly simple. A serious critique of the model could lead one rapidly into some intricate issues. For instance, is the implicit reference to a single "precinct net party strength" dimension truly compelling if we are thinking about the variation in support for Kerry and Bush over all 5,384 precincts (spread over 47 counties) being considered?

Even if we accept the simple framework of the model, do the results constitute a convincing case that there was not significant fraud? For instance, Mark Lindeman took the data used to produce the estimates reported in Table 2 and applied the following algorithm: "in approximately 10% of precincts, switch some uniform % of votes between 10% and 20% (or half of Kerry's share, whichever is less) from Kerry to Bush" (Lindeman 2006). Using such manipulated data to estimate the model, Lindeman reports obtaining results very similar to the ones reported in Table 2: "no change in slope (of course the intercept decreased), essentially no change in the number of zero weights (18), a substantial increase in sigma" (2006). Leaving aside the important question of whether it was feasible in Ohio to switch a fraction of the votes in a random sample of precincts selected from across the state, the implication remains that there are conceivable patterns of fraud that the approach used by Mebane and Herron (2005), based on estimating equation (1), would fail to detect.

A pattern of vote manipulation such as Lindeman imagines may not be detectable by any kind of statistical analysis, but still it is worthwhile to see whether the 2BL test builds confidence in Mebane and Herron's (2005) analysis or adds to skepticism about it.

To compute the 2BL test, I use the data collected as part of the DNC study for all Ohio precincts. To enhance comparability with the data analyzed by Mebane and Herron (2005), I exclude separately reported absentee vote counts. I compute the 2BL test for the votes recorded for Bush and for Kerry. For 88 Ohio counties and two candidates we have $T = 176$ separate test statistics, which implies an FDR-controlled critical value for the 2BL statistic of 31.1.

Now, at last, we find a 2BL test statistic that is larger than the FDR-controlled critical value. Of the 176 statistics, one is greater than 31.1. This is the statistic for the vote for Kerry in Summit County, which is $X^2_{B_2} = 42.7$. The next largest value is $X^2_{B_2} = 25.2$, for the vote for Kerry in Scioto county. In all there are 21 statistics larger than the single-test critical value of 16.9—nine statistics are for votes recorded for Bush and twelve are for votes recorded for Kerry. Three counties have statistics greater than 16.9 for both candidates' votes, namely, Cuyahoga, Paulding and Summit counties.

These results do not in a strict sense call into question the conclusions Mebane and Herron (2005) reach, at least as far as the analysis based on estimating equation (1) is concerned. Summit is not one of the counties that had constant precinct boundaries from 2002 to 2004, so precincts from Summit County were not included in the collection of precincts used to estimate equation (1). But the high proportion of the statistics that are greater than the critical value for a single test may indicate that there was vote manipulation that the earlier analysis failed to detect. Having set a single-test level of $\alpha = .05$, we might expect about five percent of the statistics to exceed the corresponding critical value. But about twelve percent ($21/176 = .119$) of the statistics exceed that value. Of the eighteen counties that have such a statistic, seven are not among the counties that had constant precinct boundaries. Nonetheless, finding that 13 of the 94 statistics that do come from such counties are larger than the single-test critical value is not especially reassuring. Since none of these 13 statistics is close to the FDR-controlled critical value—for 94 tests this would be 29.5—the situation with the constant-boundary counties is one where the test signal has not been turned on but it is not clear that it is firmly off.

Presidential Votes across the United States in 2000 and 2004

We have looked at 2BL test statistics from Florida in 2000 and 2004 and from Ohio in 2004, and we have found only one that is large once we take into account the fact that we are considering many such statistics. Are significant 2BL test results in general rare? If so, it might mean either that election fraud that involves manipulation of the votes is genuinely rare, or that the 2BL test is just not sufficiently sensitive. Or perhaps it is simply that despite all the controversy attending the voting in Florida and Ohio in recent elections, in fact those states are exceptional in having relatively little of the kinds of vote shifting that the 2BL test in principle is able to detect. Perhaps in other places—or in other notorious places—more large 2BL test statistics will appear.

To get some perspective on this, I analyze precinct data reporting votes for president across the U.S. in 2000 and 2004. Again I compute the 2BL test for the votes recorded for Bush and for Kerry. Precinct data are not readily available from every state, nor necessarily from every county in states for which some data are obtainable. I use data obtained from Dave Leip—for 35 states in 2000 and for 42 states in 2004—supplemented with other information.[4] I compute the 2BL test separately for the precincts in every county that has at least ten precincts. If at least ten precincts remain in the rest of a state, I also compute the test for those precincts together in one set. Except for the data from Ohio in 2004, I include any totals that are reported for absentee ballots as separate precincts. Counting the residual precincts from the small counties in each state as one county, the analysis uses data from 1,726 counties and 130,827 precincts in 2000 and from 1,743 counties and 143,889 precincts in 2004.[5] Controlling the FDR for counties over the whole country in each year—i.e., $T = 3452$ and $T = 3486$—gives FDR-controlled critical values of about 38.4.

From the boxplot display of the distribution of the 2BL test statistics shown in Figure 1, one can see that there are not many statistics as large as that global FDR-controlled critical value. Indeed, in all there are six counties that have 2BL statistics larger than 38.4: Los Angeles, CA, and Cook, IL, in both 2000 and 2004; DuPage, IL, and Hamilton, OH, in 2000; and Summit, OH, and Davis, UT, in 2004. The largest statistics in both years occur for Los Angeles: the statistic for the votes for Gore is $X^2_{B_2} = 54.8$, and for the votes for Kerry it is $X^2_{B_2} = 70.2$. Four other counties have statistics smaller than 38.4 but large enough to be rejected by the FDR criterion ($d > 1$). These are Latah, ID, Lake, IL, Hancock, OH, (the votes for Gore) and Philadelphia, PA, (the vote for Bush) in 2000 and DuPage, IL, (the votes for Kerry) in 2004.

[4]Data for Florida in 2000 and Ohio in 2004 are as described above. For Pennsylvania in 2004 I use data obtained from the Pennsylvania State Election Commission (in a file named PA-2004G-Presidential.xls). For Nebraska in 2004 I use data from the Secretary of State. I downloaded data for Cook County, IL, in 2004 from Cook County and Chicago election board websites.

[5]The states with at least one county in the analysis in 2000 are AK, AL, AR, AZ, CA, DC, DE, FL, HI, IA, ID, IL, IN, KS, LA, ME, MI, MN, MT, NC, ND, NH, NJ, NY, OH, PA, RI, SC, SD, TN, VA, VT, WA, WI and WY. For 2004 the states with at least one county in the analysis are AK, AL, AR, AZ, CA, CO, DC, DE, FL, GA, HI, IA, ID, IL, IN, KS, LA, MD, ME, MI, MN, MO, MT, NC, ND, NE, NH, NJ, NM, NV, NY, OH, OR, PA, RI, SC, SD, TN, TX, UT, VA, VT, WA, WI and WY.

*** Figure 1 about here ***

The omnibus FDR-controlled critical value is lofty indeed. More pragmatically—but less skeptically—we might consider each county together with only the other counties in the same state. Such a perspective would be relevant, for instance, if each state's election officials were to use the 2BL test to screen the election results from their state. In this case it is reasonable to determine the FDR-controlled critical values for each state by taking into account only the number of test statistics that may be computed for that state. That is, T equals the number of counties in the state for which there is usable data, multiplied by the number of candidates for whose vote totals we are computing the test.

Table 3 shows the results of applying state-specific FDR control. The list includes a mix of urban and rural places.

*** Table 3 about here ***

In 1934, Harris presented case studies detailing election frauds in four cities: Philadelphia, Chicago, Pittsburgh and Cleveland. He wrote, "Recent investigations have brought to light election scandals in the particular cities covered, but it would be a mistake to assume that other cities are free of election frauds" (Harris 1934, 320). Nearly seventy years later, the 2BL test marks three of these cities as worrisome. Of course the county containing Chicago (Cook, IL) and its adjacent neighbors (DuPage and Lake, IL) have already been flagged as having significant 2BL test statistics even when the omnibus FDR control is imposed, as has Philadelphia in 2000. Summit, OH, is adjacent to Cleveland. Cook, DuPage and Summit appear in Table 3 both in 2000 and 2004. In 2004 the statistics for Philadelphia exceed the single-test critical value but are not larger than the FDR-controlled critical values. Philadelphia has $X^2_{B_2} = 21.8$ and $X^2_{B_2} = 23.4$ respectively for Kerry's and Bush's vote totals. Pittsburgh escapes inclusion.[6]

The 2004 result for Davis, UT, highlights the importance of knowing how reported vote totals are being produced. The precise set of choices presented to each voter can strongly affect the 2BL test. Utah allows voters to cast a "straight party" vote. A voter who does this automatically casts a vote for every candidate who is affiliated with the indicated party. For Davis County the number of straight party votes is available for each precinct.[7] For both the Republican party

[6] Allegheny County has $X^2_{B_2} = 21.2$ for the votes for Gore but 2BL statistics less than 16.9 for Bush in 2000 and for both Bush and Kerry in 2004.

[7] I downloaded the data from the website of the Davis County Clerk.

($X^2_{B_2} = 6.6$) and the Democratic party ($X^2_{B_2} = 8.1$), the 2BL test statistics for the straight party vote counts are not large. The statistic is also not large for the candidate-specific votes cast for Bush ($X^2_{B_2} = 13.2$). The 2BL statistic for the candidate-specific votes cast for Kerry just exceeds the critical value for a single test ($X^2_{B_2} = 17.0$). It is intriguing that the 2BL test statistic for the precinct sums of the straight party Democratic votes and the candidate-specific votes for Kerry ($X^2_{B_2} = 42.6$) is so much larger than the statistics for the separate components. Because the straight party and candidate-specific options reflect distinct choices each voter can make, the general analysis of the 2BL test in Mebane (2006) motivates applying the test separately to each kind of vote total. Mebane (2006) identifies mechanisms that generate 2BL-distributed vote counts for a wide range of parameters that represent the votes' behavioral context, but that analysis does not imply that sums of such counts will be 2BL-distributed when the counts come from heterogeneous contexts. Indeed, too much heterogeneity will conflict with the formal requirement that the counts are mixtures of a small number of random distributions.[8] A choice to vote the party line is on its face very different from a choice not to vote the party line but, instead, to vote separately for the party's presidential candidate. Presumably this point applies to vote counts for partisan offices from any jurisdiction that allows a straight party option, which would include the vote data from all of Utah. But the point does not apply in an entirely straightforward way. For Salt Lake County the 2BL test statistics are not large if applied to the total of all votes counted for Bush ($X^2_{B_2} = 15.3$) or for Kerry ($X^2_{B_2} = 8.9$), but the statistic for the counts of the Kerry-specific votes is large ($X^2_{B_2} = 37.2$).

The large 2BL statistics for Los Angeles County certainly catch one's attention. Subdividing the data from the county mitigates the significant results to some extent. Table 4 shows, first, that significant 2BL statistics do not occur for the Absentee Groups in the data, but very large statistics occur for the collection of all the remaining precincts in each year. Treating the nonabsentee precincts from the city of Los Angeles separately, there are significant statistics both for the Los Angeles city precincts and for the remainder of the precincts. If we go further and treat each city in the county separately, then the number of cities that have at least ten precincts is 88 in 2000 and 83 in 2004. For this analysis I group the smaller cities together and treat them

[8] If too many heterogenous counts are summed, then the law of large numbers will mean the sum has approximately a normal distribution. A normal variate does not have 2BL-distributed digits.

as a set (encompassing 293 precincts in 2000 and 308 precincts in 2004). No city has a value of $X^2_{B_2}$ larger than the FDR-controlled critical values for each respective year. In 2000, nine 2BL statistics are greater than the single-test critical value, while in 2004 20 statistics are greater. This gives proportions of nominally large values of respectively $9/178 = .05$ and $20/168 = .12$. Perhaps the proportion of .05 is reassuring for the 2000 data, but the proportion of .12 echoes the naggingly worrying proportion observed for the data from Ohio 2004.

*** Table 4 about here ***

For counties over the whole country, the frequency of large 2BL test statistics does not greatly exceed the nominally expected values. There are 234 2BL test statistics greater than the single-test critical value in 2000, and there are 226 test statistics in 2004 that are that large. These counts imply proportions of large statistics not much greater than the single-test level of $\alpha = .05$ would suggest. We have $234/3422 = .068$ in 2000 and $226/3448 = .066$ in 2004. Even more then, perhaps, does the much higher proportion of nominally large 2BL statistics found for Ohio 2004 stand out.

Discussion

The good news is that, as measured by the 2BL test, signs of election fraud in recent American presidential votes seem to be rare. Several of the places that turn up with significantly large 2BL test statistics have been notorious for a century or more. That the 2BL test finds these places suggests it is probably on to something. These results using data from actual American elections tend to reinforce the simulation results of Mebane (2006) that show the 2BL test can spot many patterns of manipulation in vote counts.

A significant 2BL test result is not in itself proof of fraud. For the vote counts from Davis, UT, the test picked up the fact that two different kinds of votes were being added together in each precinct—the vote counts were sums of straight party votes and candidate-specific votes—but once the component votes were separated, indications that the votes were manipulated mostly dissipated. The marginally significant 2BL statistic found for the candidate-specific votes for Kerry might justify additional checks to verify the vote counts' accuracy, but in light of the large number of test statistics being examined from different places, the value of $X^2_{B_2} = 17.0$ is not significant if the FDR is controlled. Indeed, taking into account only the test statistics for the straight-party Democratic votes and the votes for Kerry in Davis County, so that $T = 2$, the FDR-controlled critical value corresponding to a single-test level of $\alpha = .05$ is 19.0.

The 2BL test is strikingly insensitive to some kinds of distortions that we know affected many votes. The most interesting case here is Florida, 2000. Notwithstanding the well established fact that tens of thousands of votes were lost to undervotes and overvotes throughout the state, the 2BL test does not signal any significant problems with the precinct vote totals.

The 2BL test gives a mixed message about Ohio, 2004. We can clearly reject the hypothesis that precinct vote counts throughout the state follow the 2BL distribution. The 2BL test statistic for Summit County is significantly large even when we take the FDR fully into account. Also, suspiciously many counties have 2BL test statistics that exceed the critical value we would use if we were looking at only one test. The 2BL test results do not overturn previous judgments that manipulation of reported vote totals did not determine the election outcome in Ohio, but neither do they completely dissipate the odor of suspicion that continues to hang over the state's results.

On the whole, this look at recent presidential election results through the lens of the 2BL test enhances the case that it is worth taking seriously as a statistical test for election fraud. The 2BL test cannot detect all kinds of fraud, and significant 2BL test results may occur even when vote counts are in no way fraudulent. But, considering the results from Florida in 2000, the test seems not to be confused by some kinds of distortions in elections that do not involve manipulating the vote totals. Further investigations of the test's performance are clearly warranted.

In any case, the 2BL test on its own should not be considered proof either that election fraud has occurred or that an election was clean. A significant 2BL test result can be caused by complications other than fraud. Some kinds of fraud the 2BL test cannot detect. Indeed, in the worst imaginable case where someone is able to fake an entire set of precinct vote counts, it would not be difficult to use fraudulent counts with second digits that follow the 2BL distribution. Using simulation mechanisms similar to those introduced in Mebane (2006), it would not be difficult to produce completely faked vote counts that also satisfy additional constraints such as having the artificial vote totals match the number of voters who actually turned out to vote on election day. But the larger the number of independent constraints the counts must satisfy, the more difficult it may be to produce faked counts that can escape detection. In terms of statistical testing for fraud, this is where analysis using covariates and beliefs about relationships the vote counts should satisfy may be helpful. When suitable data are available, additional analysis using covariates, robust estimation and outlier detection (Mebane and Sekhon 2004) should be conducted. Such analysis may help to diagnose the origins and perhaps put bounds on the scope of any anomalies in the vote counts.

Ultimately, however, statistical analysis after the fact can accomplish only so much. To permit statistical testing to occur in a way that can make a difference, it is crucial that the highly disaggregated data needed to support analysis be made available in as timely a manner as possible. This includes not only the precinct level vote totals needed for the 2BL test, but also information about voting machines, poll workers, voter registration, the number of ballots cast, ballot formats and other data that other analytical methods can use. Candidates, parties, political activists, citizens and other observers may need to take steps well in advance of the election to make sure local election administrators will be able to supply the necessary data in time to inform any election contests that may occur. But to prevent election fraud, appropriate practices need to be used while the election is being conducted. Insecure or opaque voting technology or election administration procedures should not be used. The election environment should not foment chaos and confusion. Not only should elections be secure and fair, but everyone should know they are secure and fair.

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.

Canedy, Dana. 2002. "Again, Sunshine State Is in Dark a Day After the Vote." *New York Times*. September 12.

Canedy, Dana. 2003. "Order Soundly Defeats Chaos as Florida County Goes to Polls." *New York Times*. March 12.

Feldman, Ariel J., J. Alex Halderman, and Edward W. Felten. 2006. "Security Analysis of the Diebold AccuVote-TS Voting Machine." URL http://itpolicy.princeton.edu/voting (accessed September 13, 2006).

Florida Legislative Staff. 2001. "FREDS 2000 dataset, PlanStatisticsT00.zip." Accessed November 16, 2001, URL `http://www.flsenate.gov/senateredistricting/freds_data.cfm`.

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

Harris, Joseph P. 1934. *Election Administration in the United States*. Washington, DC: Brookings Institution.

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

House Judiciary Committee Democratic Staff. 2005. "Preserving Democracy: What Went Wrong in Ohio." URL http://www.house.gov/judiciary_democrats/ohiostatusrept1505.pdf.

Laughlin, Meg and David Kidwell. 2004. "No Flaw is Found in Bush's State Win." *Miami Herald*. November 28.

Lindeman, Mark. 2006. "multinomRob result." Personal communication (email), August 20, 2006.

MacManus, Susan A. 2004. "Goodbye Chads, Butterfly Ballots, Overvotes and Recount Ruckuses! Election Reform in Florida: 2000-2003." In Daniel J Palazzolo and James W Ceaser, editors, *Election Reform: Politics and Policy*. Lanham, MD: Lexington Books. Pages 377–58.

Mebane, Walter R., Jr. 2004a. "Communications regarding optical voting machines in Florida, 2004." URLs http://macht.arts.cornell.edu/wrm1/commondreams/commondreams.html and http://macht.arts.cornell.edu/wrm1/ustogether/ustogether.html.

Mebane, Walter R., Jr. 2004b. "The Wrong Man is President! Overvotes in the 2000 Presidential Election in Florida." *Perspectives on Politics* 2 (Sep.): 525–535.

Mebane, Walter R., Jr. 2006. "Election Forensics: Vote Counts and Benford's Law." Paper prepared for the 2006 Summer Meeting of the Political Methodology Society, UC-Davis, July 20–22.

Mebane, Walter R., Jr. and Michael C. Herron. 2005. "Ohio 2004 Election: Turnout, Residual Votes and Votes in Precincts and Wards." In Democratic National Committee Voting Rights Institute, editor, *Democracy at Risk: The 2004 Election in Ohio*. Washington, D.C.: Democratic National Committee.

Mebane, Walter R., Jr. and Jasjeet S. Sekhon. 2004. "Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data." *American Journal of Political Science* 48 (Apr.): 392–411.

Miller, Mark Crispin. 2005. *Fooled Again*. New York: Basic Books.

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

Roig-Franzia, Manuel and Dan Keating. 2004. "Latest Conspiracy Theory—Kerry Won—Hits the Ether." *Washington Post*. November 10.

Rubin, Aviel D. 2006. *Brave New Ballot*. New York: Morgan Road Books.

Saltman, Roy G. 2006. *The History and Politics of Voting Technology*. New York: Palgrave Macmillan.

Sekhon, Jasjeet S. 2004. "The 2004 Florida Optical Voting Machine Controversy: A Causal Analysis Using Matching." URL http://sekhon.berkeley.edu/papers/SekhonOpticalMatch.pdf.

Stolberg, Sheryl Gay and James Dao. 2005. "Congress Ratifies Bush Victory After a Rare Challenge." *New York Times*. January 7.

Voting Rights Institute, Democratic National Committee. 2005. *Democracy at Risk: The 2004 Election in Ohio*. Washington, D.C.: Democratic National Committee.

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.

Wand, Jonathan N. 2004. "Evaluating the Impact of Voting Technology on the Tabulation of Voter Preferences: The 2004 Presidential Election in Florida." URL http://wand.stanford.edu/elections/us/FL2004/WandFlorida2004.pdf.

Zetter, Kim. 2004. "Researchers: Florida Vote Fishy." *Wired*. URL http://www.wired.com/news/evote/0,2645,65757,00.html (November 18).

Table 1: Frequency of Second Digits according to Benford's Law

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	.120	.114	.109	.104	.100	.097	.093	.090	.088	.085

Table 2: Vote for Kerry versus Bush: 2002 Gubernatorial Vote Regressor

	Precincts		
Variable	Coef.	SE	t-ratio
(Intercept)	0.456	0.00589	77.5
Logit(Democratic Vote in 2002)	1.040	0.00627	166.0

	Wards		
Variable	Coef.	SE	t-ratio
(Intercept)	0.64	0.0224	28.6
Logit(Democratic Vote in 2002)	1.04	0.0266	39.1

Notes: Robust (tanh) overdispersed binomial regression estimates. For each precinct or ward, the dependent variable counts the number of votes for Kerry versus the number of votes for Bush. Precincts: LQD $\sigma = 2.98$; tanh $\sigma = 2.87$; $n = 5{,}384$; 17 outliers. Wards: LQD $\sigma = 9.09$; tanh $\sigma = 8.91$; $n = 357$; no outliers.

Table 3: Counties with Signficant 2BL Tests using State-specific FDR Control

	2000				
		Gore votes		Bush votes	
County	J	d_2	$X^2_{B_2}$	d_2	$X^2_{B_2}$
Los Angeles, CA	5,045	5,011	54.8	4,930	20.3
San Bernardino, CA	910	854	28.0	865	6.9
Kent, DE	61	61	9.0	61	22.2
Latah, ID	34	31	36.7	34	3.8
Cook, IL	5,179	5,097	46.7	4,145	24.4
Dupage, IL	714	714	28.0	714	41.6
Lake, IL	403	403	33.7	402	16.1
Passaic, NJ	295	295	27.7	294	5.6
Hamilton, OH	1,025	1,020	48.7	988	8.9
Hancock, OH	67	67	34.3	67	9.9
Summit, OH	624	624	31.6	612	11.6
Lancaster, PA	225	225	29.1	225	8.3
Philadelphia, PA	1,681	1,680	29.5	1,249	34.7
King, WA	2,683	2,665	27.0	2,641	8.9

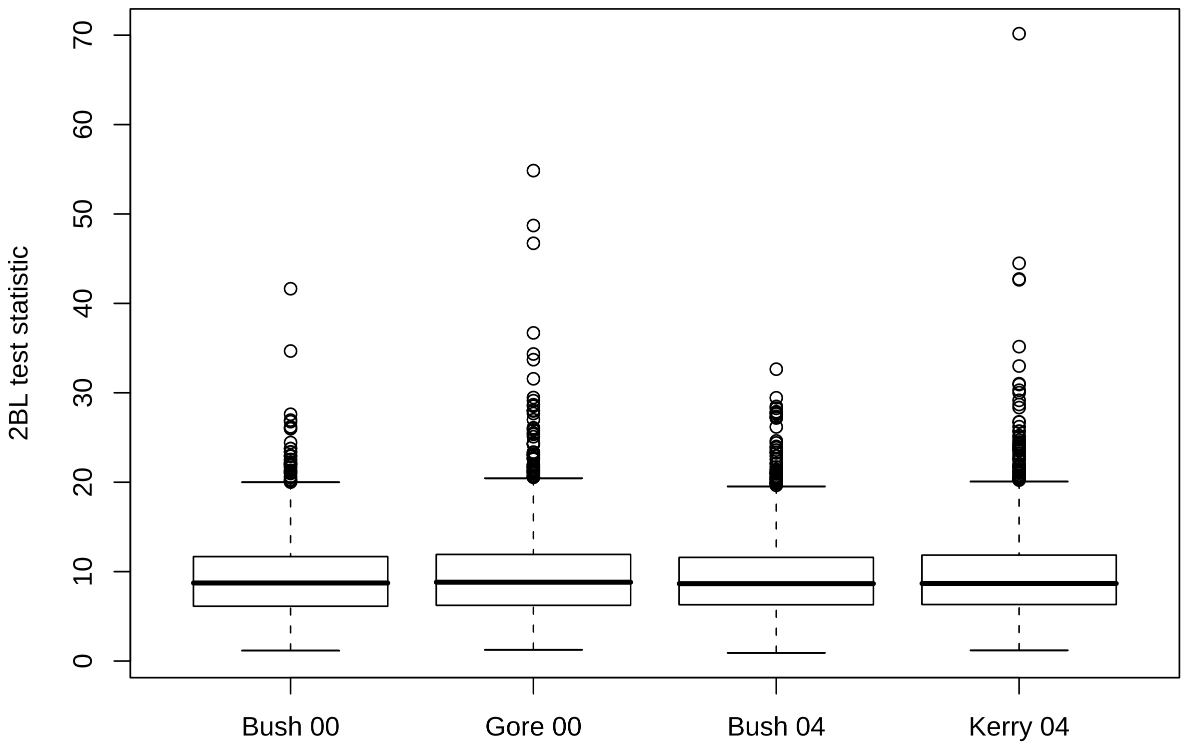
	2004				
		Kerry votes		Bush votes	
County	J	d_2	$X^2_{B_2}$	d_2	$X^2_{B_2}$
Alameda, CA	997	890	26.7	862	9.7
Glenn, CA	23	23	2.8	23	27.9
Los Angeles, CA	4,984	4,951	70.2	4,929	12.4
Nevada, CA	134	123	13.5	121	24.0
Orange, CA	1,985	1,887	26.2	1,904	32.6
San Francisco, CA	577	577	28.7	568	16.9
Jefferson, CO	324	323	30.0	323	10.4
Kootenai, ID	75	75	30.9	75	12.1
Cook, IL	4,562	4,561	44.5	4,026	27.8
DuPage, IL	732	732	35.2	732	9.1
Rock Island, IL	121	121	24.2	121	8.4
Clay, MO	76	76	28.4	76	4.0
Summit, OH	475	475	42.7	474	21.0
Davis, UT	213	212	42.6	213	6.0
Utah, UT	247	241	9.2	246	27.6
Benton, WA	177	168	29.2	173	14.8

Table 4: Selected 2BL Tests for Parts of Los Angeles County, CA

	2000				
		Gore votes		Bush votes	
County	J	d_2	$X^2_{B_2}$	d_2	$X^2_{B_2}$
Absentee Group	82	48	8.4	40	9.1
Not Absentee	4,963	4,963	55.6	4,890	21.4
Los Angeles City	1,757	1,757	27.0	1,719	19.0
not Los Angeles City	3,206	3,206	38.6	3,171	18.4

	2004				
		Kerry votes		Bush votes	
County	J	d_2	$X^2_{B_2}$	d_2	$X^2_{B_2}$
Absentee Group	382	349	11.2	328	11.3
Not Absentee	4,602	4,602	68.4	4,601	12.1
Los Angeles City	1,658	1,658	19.4	1,657	6.1
not Los Angeles City	2,944	2,944	55.0	2,944	11.2

Figure 1: Precinct-level US 2000 and 2004 Presidential Vote 2BL Test Statistics by County



Note: Each boxplot shows the distribution of the 2BL statistics for a candidate's precinct-level vote counts over the 1,726 counties in 2000 and 1,743 counties in 2004.