Political Analysis (2011) 19:269–272
doi:10.1093/pan/mpr024

Comment on "Benford's Law and the Detection of Election Fraud"

Walter R. Mebane Jr.

Departments of Political Science and Statistics, University of Michigan, 7735 Haven Hall, 505 South State Street, Ann Arbor, MI 48109-1045
e-mail: wmebane@umich.edu

"Benford's Law and the Detection of Election Fraud" raises doubts about whether a test based on the mean of the second significant digit of vote counts equals 4.187 is useful as a test for the occurrence of election fraud. The paper mistakenly associates such a test with Benford's Law, considers a simulation exercise that has no apparent relevance for any actual election, applies the test to inappropriate levels of aggregation, and ignores existing analysis of recent elections in Russia. If tests based on the second significant digit of precinct-level vote counts are diagnostic of election fraud, the tests need to use expectations that take into account the features of ordinary elections, such as strategic actions. Whether the tests are useful for detecting fraud remains an open question, but approaching this question requires an approach more nuanced and tied to careful analysis of real election data than one sees in the discussed paper.

Downloaded from http://pan.oxfordjournals.org/ at Tufts University on March 14, 2016

1 Introduction

In "Benford's Law and the Detection of Election Fraud," Deckert, Myagkov, and Ordeshook (2011) argue against so-called second-digit Benford's Law (2BL) tests, which Mebane (2006, 22) asserts "show strong promise to become a standard tool for detecting fraudulent election results." Of course, Mebane (2006, 22) goes on to say, "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 one should perhaps not expect too much from a test that has only the vote counts themselves to work with." Deckert, Myagkov, and Ordeshook (2011, 248) are not deterred by such caveats, arguing that "any inference that the analysis of official returns can begin and end with Benford's Law or that we can dispense with measuring other variables such as the socioeconomic correlates of voting is unwarranted." Mebane (2006) makes no such inference and I certainly will not do so here, but even though it is not clear at whom Deckert and colleagues' critique is aimed, I will consider whether their arguments against 2BL tests hold up. On one fundamental point, I think their argument is correct: that the mean of the second significant digits in vote counts is significantly different from 4.187 (the mean expected under Benford's Law) is not evidence that the vote counts are affected by fraud; but the way they argue for this is unsound.

The first error in Deckert, Myagkov, and Ordeshook (2011) is to tie 2BL tests strongly to Benford's Law. As they observe, citing Brady (2005) and Mebane (2006), the first significant digits in vote counts do not usually have the distribution implied by Benford's Law, and there are strong reasons to doubt that they should. If this is true, however, then Benford's Law simply does not apply to vote counts at all. The point is that if a process produces numbers that have digits that match Benford's Law, then it does so when the numbers are represented in arbitrary numerical bases (Berger 2005): there is nothing special about base 10. The "first digit" is an artifact of using base 10 arithmetic. No version of the many theorems that mathematicians have used to derive Benford's Law identifies a mechanism that produces Benford's Law patterns in numbers except for the first digit when using base 10 arithmetic. Referring instead to

Author's note: I thank Kirill Kalinin for helpful discussion.

© The Author 2011. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oup.com

statistical mixtures that can produce numbers that follow Benford-like distributions but not Benford's Law (Rodriguez 2004; Grendar, Judge, and Schechter 2007), Mebane (2010b) remarks that "it may be more precise to refer to second-digit Benford-like tests."

The fact that Benford's Law does not supply a purely mathematical argument for applying 2BL to vote counts reinforces Deckert and colleagues' position that applying 2BL to vote counts requires a theoretical basis. Purely mechanical efforts to find such a basis, such as the computational approach used in Mebane (2007), seem not to be sufficient. The analysis in Mebane (2007), for example, leaves unexplained the large and persistent deviation from expectations in 2000 and 2004 in votes for president cast in Los Angeles, California, that was reported in Mebane (2008). But no reasonable person believes that there was massive election fraud there that no one happened to notice in those years.

Mebane (2010b) suggests that a theoretical basis for 2BL tests may possibly be found in the strategies voters use in elections, citing in particular strategies discussed by Cox (1994) and Alesina and Rosenthal (1995) in connection with a number of American elections. Deckert, Myagkov, and Ordeshook (2011, 245) mention the possibility that "voters are sophisticated and strategic" and cite Cox, but they do not follow up on the implications of this idea. The point is that the second significant digits of vote counts may have expectations in normal elections that differ from 4.187 because of the political environment and voters' strategies. Fraud would be indicated by deviation from these contingent expectations and not always by deviations from 4.187. Such arguments can explain the deviations from 4.187 seen not only in presidential votes in Los Angeles in 2000 and 2004 but also in many other elections in many other years in the United States (Mebane 2010b).

The premium in trying to develop such theory is to make sure that one is focusing on real elections. In this light, the simulation exercise Deckert, Myagkov, and Ordeshook (2011) discuss is irrelevant to the task of deciding whether 2BL tests are worthwhile. Two points come immediately to mind when reading about their simulation. First is that there is no need to run the simulation to know it will show that 2BL tests work poorly. They use normal distributions for the random components in their simulated vote counts, but the second significant digits of numbers that follow a normal distribution in general do not satisfy Benford's Law. So by design, their tests that expect second-digit means to equal 4.187 will fail. The second point is that there is no reason to believe that the simulated data resemble the data from any actual election. Deckert, Myagkov, and Ordeshook (2011, 249) "begin then with the usual Downsian model of an electorate in which eligible voters occupy positions in a two-dimensional space, where, if they vote, they do so for the candidate closest to their ideal position." It is pretty clear that a two-dimensional spatial model does not describe actual behavior in any electorate; precisely, the authors fail to show that the digits from such a simulation match the digits in any actual election. It is hard to know what we may learn from such an exercise.

The challenge taken up by Mebane (2006) is exactly the opposite of what Deckert and colleagues do. Given real election data (from Florida) in which the second significant digits match 2BL (according to chi-squared tests), Mebane (2006) finds two simulation mechanisms that match the real-world data. Moreover, in data from Mexico, Mebane (2006) finds a mechanism that matches the difference between the digits in vote counts at different levels of aggregation (ballot boxes [*casillas*] versus precincts [*secciones*]). The question then becomes whether these real-world patterns generalize—and what precisely is the pattern whose generalizability one is investigating. Mebane pursues this question in subsequent work (e.g., Mebane 2007, 2010a).

An issue in both the simulation exercise and in Deckert and colleagues' application of digit tests to data from Ukraine is that Deckert and colleagues use 2BL tests with vote counts at levels of aggregation much higher than the precinct-level data that Mebane investigates. Even though Deckert, Myagkov, and Ordeshook (2011, 251) simulate what they refer to as "precincts" with 10,000 and 20,000 simulated voters, such large sizes are never found in real precinct-level data. The data they use from Ukraine are for *rayons* (counties). Their motivation for using 2BL tests with data at such a high level of aggregation is unclear.

The last set of issues with Deckert, Myagkov, and Ordeshook (2011) concerns their treatment of Russia. Citing Myagkov, Ordeshook, and Shaikin (2009), Deckert and colleagues emphasize fraudulent patterns in elections since 2004 especially in Russia's ethnic republics, with a key point being that there are places where turnout was at or near 100% with the party of power receiving nearly 100% of the votes. But Mebane and Kalinin (2009a, 2009b) show that the last digits of precinct vote totals are too often zero or five and that precinct turnout figures appear with round numbers above 60%—specifically there are high

Downloaded from http://pan.oxfordjournals.org/ at Tufts University on March 14, 2016

frequencies at 60%, 70%, 80%, 90%, and 100% turnout—and such fraudulent patterns occur perhaps more dramatically in oblasts than in republics. Kalinin and Mebane (2011) argue that these rounded figures (and rounded figures in territory-level statistics) are part of a system in which regional governors signal loyalty to the center; Kalinin and Mebane (2011) show that the occurrence of rounded figures is related to financial transfers from the center to regions in ways the signaling argument implies. Deckert, Myagkov, and Ordeshook (2011) may perhaps be forgiven for overlooking this work, which came into existence after they had completed their paper. But they also ignore Mebane and Kalinin (2010), which uses 2BL tests to try to localize and diagnose both strategy and fraud in the 2007 and 2008 elections. The analysis in Mebane and Kalinin (2010) uses 2BL tests in conjunction with a range of covariates and so may not fall under Deckert and colleagues' critique of a "universally applicable magic box into which we plug election statistics and out of which comes an assessment of an election's legitimacy." I do not mean to assert here that the analysis in Mebane and Kalinin (2010) is correct—its brevity marks it as clearly preliminary, and even taking that into account it may be wrong—but it demonstrates that 2BL tests can be used in a context that takes seriously the fact that "fraud is not uniformly distributed across the country" (Deckert, Myagkov, and Ordeshook 2011, 257).

Deckert, Myagkov, and Ordeshook (2011, 246) complain about "research" on "the application of Benford's Law to elections" being "anything but peer reviewed." That is a fair complaint, but they have taken on the mission of writing "an assessment of the conclusions that might apply were peer review a part of this component of the discipline's public persona" by referring to several conference papers. With any luck, work will soon appear that meets the standard of peer review and also offers some theory to guide the use of 2BL tests, perhaps along the lines mentioned above that draws in strategic behavior. If Deckert, Myagkov, and Ordeshook (2011) would clarify what is the target of the purely negative contribution they make in the piece discussed here, perhaps it would be possible to tell whether it is worthwhile. For instance, do they mean to call into question the conclusions reached in Mebane (2010c), based partly on 2BL tests, about fraud in the Iranian 2009 election? As I said at the beginning, I do not dispute that 2BL tests need a foundation in some kind of theory to be useful. Deckert, Myagkov, and Ordeshook (2011) offer no such theory, nor any hint of how such theory might come to be.

References

Alesina, Alberto, and Howard Rosenthal. 1995. *Partisan politics, divided government, and the economy*. New York, NY: Cambridge University Press.

Berger, Arno. 2005. Multi-dimensional dynamical systems and Benford's Law. *Discrete and Continuous Dynamical Systems* 13: 219–37.

Brady, Henry E. 2005. Comments on Benford's Law and the Venezuelan election Working paper, January 19, 2005.

Cox, Gary W. 1994. Strategic voting equilibria under the single nontransferable vote. *American Political Science Review* 88: 608–21.

Deckert, Joseph, Mikhail Myagkov, and Peter C. Ordeshook. 2011. Benford's Law and the detection of election fraud. *Political Analysis* 19(3):245–68.

Grendar, Marian, George Judge, and Laura Schechter. 2007. An empirical non-parametric likelihood family of data-based Benford-like distributions. *Physica A* 380:429–38.

Kalinin, Kirill, and Walter R. Mebane, Jr. 2011. Understanding electoral frauds through evolution of Russian federalism: From 'Bargaining Loyalty' to 'Signaling Loyalty'. Paper prepared for the 2011 Annual Meeting of the Midwest Political Science Association, Chicago, IL, March 31–April 2.

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, Davis, CA, July 20–22.

———. 2007. Statistics for digits. Paper presented at the 2007 Summer Meeting of the Political Methodology Society, Pennsylvania State University, State College, PA, July 18–21.

———. 2008. Election forensics: The second-digit Benford's law test and recent American presidential elections. In *Election Fraud: Detecting and Deterring Electoral Manipulation*, ed. R. Michael Alvarez, Thad E. Hall, and Susan D. Hyde. Washington, DC: Brookings Institution.

———. 2010a. Election fraud or strategic voting? Paper prepared for the 2010 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 22–25.

———. 2010b. Election fraud or strategic voting? Can second-digit tests tell the difference? Paper prepared for the 2010 Summer Meeting of the Political Methodology Society, University of Iowa, Iowa City, IA, July 22–24.

———. 2010c. Fraud in the 2009 presidential election in Iran? *Chance* 23:6–15.

Mebane, Walter R., Jr., and Kirill Kalinin. 2009a. Comparative election fraud detection. Paper prepared for the 2009 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 2–5.

Downloaded from http://pan.oxfordjournals.org/ at Tufts University on March 14, 2016

———. 2009b. Electoral falsification in Russia: Complex diagnostics selections 2003-2004, 2007-2008 (in Russian). *Electoral Policy* REO 2/09: 57–70.

———. 2010. Electoral fraud in Russia: Vote counts analysis using second-digit mean tests. Paper prepared for the 2010 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 22–25.

Myagkov, Mikhail, Peter C. Ordeshook, and Dimitry Shaikin. 2009. *The forensics of election fraud: With applications to Russia and Ukraine*. New York, NY: Cambridge University Press.

Rodriguez, Ricardo J. 2004. First significant digit patterns from mixtures of uniform distributions. *American Statistician* 58: 64–71.

Downloaded from http://pan.oxfordjournals.org/ at Tufts University on March 14, 2016