The Devil Is in the Digits[1]

By Bernd Beber and Alexandra Scacco
Saturday, June 20, 2009 12:02 AM

Since the declaration of Mahmoud Ahmadinejad's landslide victory in Iran's presidential election, accusations of fraud have swelled. Against expectations from pollsters and pundits alike, Ahmadinejad did surprisingly well in urban areas, including Tehran -- where he is thought to be highly unpopular -- and even Tabriz, the capital city of opposition candidate Mir Hussein Mousavi's native East Azarbaijan province.[2]

Others have pointed to the surprisingly poor performance of Mehdi Karroubi, another reform candidate, and particularly in his home province of Lorestan, where conservative candidates fared poorly in 2005, but where Ahmadinejad allegedly captured 71 percent of the vote.[3] Eyebrows have been raised further by the relative consistency in Ahmadinejad's vote share across Iran's provinces, in spite of wide provincial variation in past elections.[4]

These pieces of the story point in the direction of fraud, to be sure. They have led experts to speculate that the election results released by Iran's Ministry of the Interior had been altered behind closed doors. But we don't have to rely on suggestive evidence alone. We can use statistics more systematically to show that this is likely what happened. Here's how.

We'll concentrate on vote counts -- the number of votes received by different candidates in different provinces -- and in particular the last and second-to-last digits of these numbers.[5] For example, if a candidate received 14,579 votes in a province (Mr. Karroubi's actual vote count in Isfahan), we'll focus on digits 7 and 9.

This may seem strange, because these digits usually don't change who wins.

---

[1] This version of the annotated op-ed was updated on 25 June 2009. The previous version of this document is archived at http://www.columbia.edu/~bhb2102/files/Beber_Scacco_The_Devil_Is_in_the_Digits_v3.pdf.

[2] Matt Smith, "Analysts pore over 'ambiguous' results in Iran," 16 June 2009, available at http://www.cnn.com/2009/WORLD/meast/06/16/iran.election.questions/index.html.

[3] Nate Silver, "If He Did It," 16 June 2009, http://www.fivethirtyeight.com/2009/06/if-he-did-it.html.

[4] Juan Cole, "Stealing the Iranian Election," 13 June 2009, http://www.juancole.com/2009/06/stealing-iranian-election.html.

[5] For an analysis of first digits in Iran's 2009 election, see Boudewijn F. Roukema, "Benford's Law Anomalies in the 2009 Iranian Presidential Elections," http://arxiv.org/abs/0906.2789v1. For an analysis of second digits, see Walter R. Mebane, "Note on the presidential election in Iran, June 2009," http://www-personal.umich.edu/~wmebane/note18jun2009.pdf.

In fact, last digits in a fair election don't tell us anything about the candidates, the make-up of the electorate or the context of the election.

They are random noise in the sense that a fair vote count is as likely to end in 1 as it is to end in 2, 3, 4, or any other numeral.[6] But that's exactly why they can serve as a litmus test for election fraud. For example, an election in which a majority of provincial vote counts ended in 5 would surely raise red flags.

Why would fraudulent numbers look any different? The reason is that humans are bad at making up numbers. Cognitive psychologists have found that study participants in lab experiments asked to write sequences of random digits will tend to select some digits more frequently than others.[7]

So what can we make of Iran's election results? We used the results released by the Ministry of the Interior and published on the web site of Press TV, a news channel funded by Iran's government.[8] The ministry provided data for 29 provinces, and we examined the number of votes each of the four main candidates -- Ahmadinejad, Mousavi, Karroubi and Mohsen Rezai -- is reported to have received in each of the provinces -- a total of 116 numbers.

The numbers look suspicious. We find too many 7s and not enough 5s in the last digit.[9] We expect each digit (0, 1, 2, and so on) to appear at the end of 10 percent of the vote counts. But in Iran's provincial results, the digit 7 appears 17 percent of the time, and only 4 percent of the results end in the number 5. Two such departures from the average -- a spike of 17 percent or more in one digit and a drop to 4 percent or less in another --

[6] We prove in a working paper that last digits will be distributed uniformly given weak assumptions ("What the Numbers Say: A Digit-Based Test for Election Fraud Using New Data from Nigeria", which is available at http://www.columbia.edu/~bhb2102/files/Beber_Scacco_ElectionFraud.pdf). This working paper also presents data from Sweden's 2002 parliamentary election to show that last digits occur with equal frequency in fair elections.

[7] Philip J. Boland and Kevin Hutchinson, Student selection of random digits, Statistician, 49(4): 519-529, 2000. Alphonse Chapanis, Random-number guessing behavior, American Psychologist, 8: 332, 1953. Gustave J. Rath, Randomization by humans, American Journal of Psychology, 79(1): 979, 1966. For a literature review, see Raymond S. Nickerson, Perception and production of randomness, Psychological Review, 109(2): 330-357, 2002.

[8] Press TV, "Interior Ministry releases provincial vote count", 15 June 2009, available at http://www.presstv.ir/detail.aspx?id=98177&sectionid=351020101. A copy of the data is available at http://www.columbia.edu/~bhb2102/files/Iran_2009.csv.

[9] R code for the complete analysis is located at http://www.columbia.edu/~bhb2102/files/Iran_2009.R. In analysis that preceded Iran's election, we found that our tests for (a) high levels of variation in digit frequencies and (b) a lack of non-adjacent digit pairs were most sensitive to fraud in the case of Nigeria's 2003 elections. This is why we ran these, and not other, tests on the data from Iran. Details about the earlier analysis can be found in the working paper referenced in fn. 6.

are extremely unlikely.[10] Fewer than four in a hundred non-fraudulent elections would produce such numbers.[11]

As a point of comparison, we can analyze the state-by-state vote counts for John McCain and Barack Obama in last year's U.S. presidential election.[12] The frequencies of last digits in these election returns never rise above 14 percent or fall below 6 percent, a pattern we would expect to see in seventy out of a hundred fair elections.

But that's not all. Psychologists have also found that humans have trouble generating non-adjacent digits (such as 64 or 17, as opposed to 23) as frequently as one would expect in a sequence of random numbers.[13] To check for deviations of this type, we examined the pairs of last and second-to-last digits in Iran's vote counts. On average, if the results had not been manipulated, 70 percent of these pairs should consist of distinct, non-adjacent digits.[14]

Not so in the data from Iran: Only 62 percent of the pairs contain non-adjacent digits. This may not sound so different from 70 percent, but the probability that a fair election would produce a difference this large is less than 4.2 percent. And while our first test -- variation in last-digit frequencies -- suggests that Rezai's vote counts are the most

---

[10] A different measure of variation in digit frequencies is their standard deviation, which is .039 in the case of Iran's provincial-level vote counts. (In expectation, the standard deviation is asymptotically 0.) The probability of a fair election with 116 reported vote counts producing a standard deviation of this size is .077. The same p-value is produced by a chi-square test, another commonly used hypothesis test. Using either of these latter tests has the advantage that we are evaluating a single hypothesis regarding digit frequencies as opposed to one for each distinct numeral. It also has the advantage that it maps closely onto our expectation of excess variation in last-digit frequencies in a fraudulent election, while we do not have clear expectations about which digits in particular should occur relatively frequently. These latter tests have the disadvantage of being somewhat less intuitive, because they examine measures of the variability in digit frequencies rather than the digit frequencies themselves. Note also that our strong conclusion derives from combining our test of last-digit frequencies with a test for non-adjacency in last and second-to-last digits. This combined test produces a robustly significant result that a fair election is unlikely to produce the kind of data we observe in Iran, even if we use the alternative tests described in this footnote. See fn. 15.

[11] We compute probabilities by simulating last digits for 100,000 unbiased election returns for Iran's provinces and candidates (i.e. we repeatedly draw a sample of size 116 from the set containing numerals 0 through 9, with replacement), and then checking how often these clean numbers produce the types of suspicious patterns we see in the data from Iran.

[12] The data file we used is available at http://www.columbia.edu/~bhb2102/files/US_2008.csv.

[13] See references in fn. 7, as well as F. Michael Rabinowitz, Characteristic Sequential Dependencies in Multiple-Choice Situations, Psychological Bulletin, 74(2): 141-148, 1970.

[14] This is a corollary of the fact that last digits should occur with equal frequency. For an arbitrary second-to-last numeral, there are seven out of ten equally likely last digits that will produce a non-adjacent pair. Note that we treat both 09 and 10 as adjacent.

irregular, the lack of non-adjacent digits is most striking in the results reported for Ahmadinejad.

Each of these two tests provides strong evidence that the numbers released by Iran's Ministry of the Interior were manipulated. But taken together, they leave very little room for reasonable doubt. The probability that a fair election would produce both too few non-adjacent digits and the suspicious deviations in last-digit frequencies described earlier is less than .005.[15, 16] In other words, a bet that the numbers are clean is a one in two-hundred long shot.[17]

Bernd Beber and Alexandra Scacco, Ph.D. candidates in political science at Columbia University, will be assistant professors in New York University's Wilf Family Department of Politics this fall.

---

[15] This probability is in fact .0014, quite a bit lower than .005. Since the last and second-to-last digits are independent (and hence deviations in last-digit frequencies on the one hand and deviations in non-adjacency with the second-to-last digit on the other hand are independent events), the joint probability is the product of the marginal probabilities, which are .035 (reported as less than 4%) and .041 (reported as less than 4.2%). More specifically, the probability is .0014 that a fair election (with 116 vote counts) has the characteristics that (a) 62% or fewer of last and second-to-last digits are non-adjacent, and (b) has at least one numeral occurring in 17% or more of last digits and another numeral occurring in 4% or fewer of last digits. The probability is .003 that a fair election would have the features that (a) 62% or fewer of last and second-to-last digits are non-adjacent, and (b) last-digit frequencies have a standard deviation of .039 or more.

[16] Several comments we have received point to the fact that if we include counts of spoiled votes in our analysis, our tests find no significant evidence of vote manipulation. (A chi-square test of deviations in last-digit frequencies yields a p-value of about .8, while the probability of a fair election producing as few non-adjacent digit pairs as observed in this data is about .1.) It was and is not obvious to us why there would be an incentive to fabricate counts of spoiled votes, which is why we did not think to include those counts in any of our original analysis. On the other hand, some readers have suggested, quite reasonably so, that if someone made up some province-level results wholesale, it is not clear why that person would use actual counts of spoiled votes and stop short of fabricating those as well.

Note that from a technical standpoint, we do find a statistically significant result even after adjusting for the possibility of testing our hypotheses across multiple subsets of the available province-level counts. One could argue that we should have run our tests on the vote counts of the four candidates, as we did; all vote counts, including spoiled ballots; and each of the five columns of data separately. This amounts to seven hypothesis tests, so if we apply the Bonferroni correction, we require a p-value of .007 or less for a significant result at the 95% level. As described in fn. 15, we calculate a p-value of .003 in our test of the standard deviation in last-digit frequencies and the lack of non-adjacency in last and second-to-last digits in the data from Iran. In other words, the evidence we find in our test of the four candidates' vote counts in particular is so strong that it offers significant support for our conclusions overall, even if our tests find no reason to be suspicious in a different subset of the data.

[17] The last sentence should read: "In other words, a bet that a clean election would produce these numbers is a one in two-hundred long shot." Thanks to the many readers who alerted us to this error.