What the Numbers Say: A Digit-Based Test for Election Fraud

Bernd Beber and Alexandra Scacco

Department of Politics, New York University, 19 West 4th Street, New York, NY 10012
e-mail: bernd.beber@nyu.edu (corresponding author), alex.scacco@nyu.edu

Edited by R. Michael Alvarez

Is it possible to detect manipulation by looking only at electoral returns? Drawing on work in psychology, we exploit individuals' biases in generating numbers to highlight suspicious digit patterns in reported vote counts. First, we show that fair election procedures produce returns where last digits occur with equal frequency, but laboratory experiments indicate that individuals tend to favor some numerals over others, even when subjects have incentives to properly randomize. Second, individuals underestimate the likelihood of digit repetition in sequences of random integers, so we should observe relatively few instances of repeated numbers in manipulated vote tallies. Third, laboratory experiments demonstrate a preference for pairs of adjacent digits, which suggests that such pairs should be abundant on fraudulent return sheets. Fourth, subjects avoid pairs of distant numerals, so those should appear with lower frequency on tainted returns. We test for deviations in digit patterns using data from Sweden's 2002 parliamentary elections, Senegal's 2000 and 2007 presidential elections, and previously unavailable results from Nigeria's 2003 presidential election. In line with observers' expectations, we find substantial evidence that manipulation occurred in Nigeria as well as in Senegal in 2007.

1 Introduction

Suppose you have been asked to assess how ''clean'' a past national election was in different areas of a country. You have poor national-level information about the make-up of the voting population and virtually no information at the subnational level. You do not have access to results from any previous elections. Constituency maps are either not publicly available or do not exist. In essence, the only information you have is a list of electoral returns. This is a situation election monitors are likely to encounter in a range of developing countries where fraud may occur, and it is a situation in which regression-based tests for outliers will not work. Is it possible to say, with some confidence, whether results have been manipulated by looking at the return sheets only?

We approach this problem by developing a digit-based test that exploits human biases in number generation.[1] We refer to four expectations from the relevant psychology literature: First, fair election procedures should produce returns where last digits occur with equal frequency, but laboratory experiments have shown that individuals tend to disproportionately select particular numerals, even when they have incentives to properly randomize. Second, individuals tend to underestimate the likelihood of digit repetition in sequences of random integers, which means that we should observe relatively fewer instances of repeated numbers on manipulated vote report sheets. Third, laboratory experiments demonstrate a preference for pairs of adjacent digits, which suggests that such pairs should be abundant on fraudulent return sheets. Fourth, subjects avoid pairs of distant numerals (i.e., digits that are neither repetitive nor adjacent), so those should appear with lower frequency on manipulated returns.

Authors' note: Supplementary materials for this article are available on the *Political Analysis* Web site.
[1]In the course of the review process, we became aware of a set of papers that independently developed a related test in the field of research ethics (Mosimann, Wiseman, and Edelman 1995; Mosimann and Ratnaparkhi 1996; Mosimann et al. 2002). We provide alternative theoretical foundations, a set of tests for trailing digit pairs as opposed to terminal digits alone and novel data. Additional information is available in the supplementary materials posted on the *Political Analysis* Web site.

Our approach to detecting electoral fraud is similar to work by Mebane (2006, 2008) in that we high-light digit patterns in reported vote counts. In contrast to Mebane's inspection of second digits, we focus on last digits and digit pairs, which permits us to be relatively agnostic with respect to the underlying distribution from which vote counts are drawn.[2] Although it may well be reasonable to assume that the second digits of electoral returns follow a Benford distribution, as Mebane suggests, we show that our last-digit test requires extremely weak distributional assumptions.[3]

The digit-oriented procedure we propose is different from but complementary to other forensic methods used to detect election fraud. Aside from digit-based methods, two other types of approaches feature prominently (Alvarez, Hall, and Hyde 2008):[4] First, some authors use econometric tools to identify suspicious variation in vote counts. Such variation can be temporal, as in analyses of vote flows between candidates from one election to the next, or spatial, as in analyses of deviations from predictions based on districts' socioeconomic profiles or other characteristics.[5] Second, some authors analyze turnout rates and suggest that high levels and peculiar distributions of turnout across districts can be indicative of electoral irregularities (Myagkov, Ordeshook, and Shakin 2005, 2007, 2009). Myagkov et al. also argue that the relationship between turnout rates and vote counts should be reasonable in fair elections, for example, in the sense that an increase in turnout does not lead to a decrease in the overall number of votes for a given candidate.

These alternative techniques work best if the available data go beyond merely an election's vote counts and include information on constituency characteristics, returns from previous elections, or the size of the eligible electorate for turnout measures. The method we develop here can be deployed in data-rich environments, but it can be applied to cases where data availability is poor too, and these may well be cases in which fraud is particularly likely to occur in the first place. We also propose a method that is sensitive to a particular variety of fraud: A test based on turnout data may register stuffed ballot boxes and intimidated voters, whereas our tests only detect fabricated vote counts at the tabulation stage.[6] For this reason, the tests we develop here should be seen as a complement, not a replacement of existing techniques of fraud detection.

We proceed as follows: Section 2 shows that last digits will occur with equal frequency for a large class of theoretical distributions, and we argue that nonfraudulent electoral returns are likely drawn from such a distribution. We also derive implications for pairs of last and penultimate digits. The section then presents simulations supporting our claim that the last digits of electoral results are likely to be distributed uniformly, provided two conditions are met: Counts do not cluster within a very narrow band of numbers, for example, because election units serve the same number of people with the same preferences over parties or candidates, and counts do not include a large share of very small numbers. The section concludes by arguing that the last digits of fabricated counts are not necessarily uniformly distributed. Section 3 provides an answer as to why this is the case and discusses common human biases in random number generation. Section 4 presents empirical evidence using data from two elections where compelling allegations of fraud have been made (Nigeria's 2003 and Senegal's 2007 presidential elections) and two elections where suspicions of fraud have been muted or entirely absent (Sweden's 2002 parliamentary and Senegal's 2000 presidential elections). Section 5 concludes.

2   Theoretical Expectations for Digit Distributions

Before we can assess the extent to which observed digit distributions on electoral return sheets suggest manipulation, we need to establish baseline expectations about how digits should be distributed in a fair election. In this section, we first show that discrete distributions with certain characteristics yield (1) last

---

[2]Cantú and Saiegh (2011) propose Benford-distributed leading digits and develop a machine-learning method that uses simulated data, calibrated against authentic electoral data, to classify electoral contests during Argentina's "infamous decade" from 1931 to 1941 as fraudulent or clean.

[3]Nigrini (1999) applies Benford's law to detect tax and accounting fraud, and Schäfer et al. (2004) use it to identify fake surveys. See also the critique by Deckert, Myagkov, and Ordeshook (2011) and the response by Mebane (2011).

[4]Levin et al. (2009) apply all three kinds of tests to data from Venezuela. Myagkov, Ordeshook, and Shakin (2007, 2009) provide detailed analyses of data from Russia and Ukraine.

[5]See, for example, Mebane and Sekhon (2004) and Wand et al. (2001), who use a number of different methods.

[6]In particular, our tests can detect numbers that have been fabricated by officials without the aid of a count-generating algorithm.

digits that are distributed uniformly (e.g., the number 2 is as likely to appear as the number 5), and (2) sequences where the probabilities with which different numbers appear after a given digit are uniformly distributed as well (e.g., the number 2 is as likely to appear as the number 5 after the number 1). As a general rule, this result obtains if a distribution meets two requirements: First, the density in its tails approaches zero, and second, it can be approximated in a piecewise linear fashion over intervals of a size corresponding to the base of the numeral system. We state and prove these requirements more formally below.[7] We then show, by way of simulations, that a wide variety of theoretical distributions meet these requirements.

2.1 *Conditions Under Which Last Digits Are Uniformly Distributed*

We conceptualize the numbers observed on an electoral return sheet as draws from a random variable X, which follows some probability distribution f. For convenience, we sometimes conceive of f as a discretization of a continuous probability density function g, where

$$f(z) = \int_{z}^{z+1} g(x)\mathrm{d}x, \tag{1}$$

with $z \in \mathbb{Z}^*$.

We prove our results in general for any numeral system with base $b > 1$. The electoral returns we check for manipulation are written in base 10, as one would expect.

Proposition 1 states that the last digits of the draws from X are distributed uniformly if three conditions are met: The support of g is divisible by base b of the numeral system, the density of g at the lower bound of its support is equal to its density at the upper bound, and we can linearly approximate g in segments of size b. Here, we assume that we can approximate g without error, an assumption relaxed by proposition 2. Proposition 3 relaxes the assumption that the support of the probability density function is divisible by b, although at the cost of requiring that its density approach 0 in the tails.

Proposition 1. Consider a discrete, nonnegative random variable X with probability density f, where f discretizes continuous probability density function g by piecewise integration over intervals of size 1, and g has domain $[s_1, s_2]$, where $s_2 - s_1$ is divisible by base b of a given positional numeral system. Then the occurrence of numerals in the last digit of X is distributed uniformly if g can be estimated by piecewise linear approximation over consecutive intervals of size b, and $g(s_1) = g(s_2)$.

All proofs are provided in the supplementary materials available on the *Political Analysis* Web site, unless stated otherwise. The intuition behind the proof for proposition 1 is that if density function g takes on the same value at its lower and upper bound, then any linear change in g over a given interval of b digits (and hence any differences in the density attributed to last digits) will eventually have to be offset by the equivalent negative change in g over another interval of size b.

Proposition 2 addresses the open question of how accurate the linear approximation of g (and in turn of f) has to be in order for uniformly distributed last digits to occur. (The proof for proposition 1 is limited to the case in which there is no error in the linear approximation of density function g, i.e., g is a piecewise linear function.)

Proposition 2. Suppose the discrete, nonnegative random variable X has probability density $f(ab) + k_a d + f_\mathrm{e}(ab + d)$ with domain $[s_1, s_2]$ for any $a \in \left\{\frac{s_1}{b}, \ldots, \frac{s_2 - b}{b}\right\}$ and $d \in \{0, \ldots, b - 1\}$, given base b of the numeral system (i.e., suppose that we can decompose f into a piecewise linear component and an "error" component $f_\mathrm{e}(.)$ within each interval of size b). Then proposition 1 holds if

$$\sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} f_\mathrm{e}(ab+d_1) = \sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} f_\mathrm{e}(ab+d_2) \tag{2}$$

for any $d_1, d_2 \in \{0, \ldots, b - 1\}$.

---

[7]For alternative proofs for numbers drawn from certain continuous distributions, see Dlugosz and Müller-Funk (2009) and Mosimann and Ratnaparkhi (1996).

In other words, our result from proposition 1 obtains if a piecewise linear approximation under- and overestimates the true density to the same extent across possible last digits. The bias induced by a linear approximation in the density of different last digits has to average out over the support of the density function. One direct implication is that we can expect last digits to appear with equal frequency if the error around the linear approximation of the density function is zero in expectation, as stated in corollary 1.

Corollary 1. Proposition 1 holds if

$$E(f_e(ab+d)) = 0 \tag{3}$$

for all $d \in \{0, \ldots, b-1\}$.

Proof: Follows directly from proposition 2.

We can provide a similar, but less strict, statement in terms of the partial derivative of the summed approximation error with respect to d. Although corollary 1 shows that last digits are uniformly distributed if the piecewise linear approximation provides on average an unbiased estimate of the density function, corollary 2 states that the approximation can in fact be biased, as long as it is biased in the same way for draws ending in different digits.

Corollary 2. Proposition 1 holds if

$$\frac{\partial \sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} f_e\left(ab+d\right)}{\partial d} = 0 \tag{4}$$

for $d \in \{0, \ldots, b-1\}$.

Proof: Follows directly from proposition 2.

It is apparent from this corollary that the key class of error functions f_e for which proposition 1 will not hold are periodic or quasi-periodic with period b. The following corollary illustrates this insight by pointing to multiplicatively or additively separable error functions: Last digits are not uniformly distributed if the error around the linear density approximation can be written as the product or sum of one function of a and b and a separate function of d.

Corollary 3. Proposition 1 does not hold if $f_e(a, b, d) = \tilde{f}_e(a, b)h_e(d)$ or $f_e(a, b, d) = \tilde{f}_e(a, b) + h_e(d)$, and $h_e(d)$ is not constant.

We will show in Section 2.2 that this is of no particular concern for a wide variety of distributions that could underpin fair electoral processes. It is difficult to think of reasons why the distribution of electoral results might follow a periodic distribution with period 10. One plausible example would be if fair election results derived from an application of a floor or ceiling function to vote returns, but we are unaware of election procedures that call for the rounding or truncation of vote counts.[8] In general, only very particular (and peculiar) distributional assumptions will consistently produce last digits that are not uniformly distributed.

Proposition 3 relaxes the assumption that the support of the distribution of electoral returns has to be a multiple of numeral base b, but in turn it requires that the density has to approach 0 for very large or very small returns. In reality, this is an issue only for small returns because we cannot reasonably extend the support of f below its natural lower bound of 0. If there is a nontrivial probability of observing less than 18 votes for a unit of interest, then proposition 3 does not hold (although proposition 1 still may).

For convenience, proposition 3 explicitly imposes a restriction on the linear approximation error, which is equivalent to the restriction discussed in corollary 1.[9]

Proposition 3. Consider a discrete, nonnegative random variable X with probability density function f and domain $\{s_1, \ldots, s_2\}$. Suppose f can be approximated by an arithmetic progression for any sequence containing $2b-1$ elements, where b is the base of the positional numeral system, and the approximation error follows function f_e, where $E[f_e(z+d)] = 0$ over $z \in \{s_1, \ldots, s_2 - 2(b-1)\}$ for any $d \in \{0, \ldots, b-1\}$.

---

[8]Of course election officials may round returns in any case, in violation of proper procedure. We consider this possibility in depth in Section 4.

[9]Note that in this case f is linearly approximated over sequences of size $2(b-1)$ rather than b, which means that proposition 3 actually places a somewhat stricter restriction on the approximation error than corollary 1.

Then the occurrence of numerals in the last digit of X approaches a uniform distribution as $f(x)$ approaches 0 for $x \leqslant s_1 + 2b - 3$ and $x \geqslant s_2 - 2b + 3$.

The intuition behind the proof of proposition 3 is similar to the one for proposition 1. Here, we show that the total density for different last digits in sequences of size $2(b-1)$ is proportional to a constant if we can linearly approximate the density function within each sequence. In the proof of proposition 1, we broke density function g into consecutive pieces of size b. Here, the pieces are overlapping, with a sequence starting at each integer, and in turn the density function's support no longer has to be divisible by b.

Finally, no formal proof is needed to see that if last digits are independently and uniformly distributed, then (a) in expectation, no last digit will be repeated more frequently than any other in a series of N random draws, and (b) the expected number of repetitions (i.e., consecutive draws of the same last digit) is $\frac{N-1}{b}$. We argue that the type of empirical data we consider lends itself to the assumption that last digits are independently distributed. It is certainly possible that the last digit of the total number of votes cast at a polling station is correlated with the last digit of the vote count at the next polling station. But if turnout is in the several hundreds, as it is in our data, it would take a spatial correlation of unlikely magnitude to carry through to the last digit.

Also note that if last and penultimate digits are independently distributed, and last digits are distributed uniformly, then the expected number of pairs with digit repetition is again $\frac{N-1}{b}$, regardless of how the penultimate digit is distributed. Even if the second-to-last digit were always the same, it would not change the fact that the last digit is a match with probability $\frac{1}{b}$. If we think about the minimum distance between penultimate and last digits more generally (for convenience, we like to visualize numerals in a circle, in which case it is easy to see that the minimum distance between 7 and 1, for example, is 4), we can say that this distance is 0 with probability $\frac{1}{b}$, it is 1 with probability $\frac{2}{b}$, and it is greater than 1 with probability $\frac{b-3}{b}$ (for $b > 2$). We later use simulations to construct confidence intervals around these expected values, and we investigate the extent to which return sheets significantly deviate from them.

2.2 *Distributions for Which Last Digits Are Uniformly Distributed*

The previous section suggested that we require very particular distributional assumptions in order for last digits not to be distributed uniformly. Figure 1 illustrates this point. We parameterize eight different distributions in terms of their first two moments and simulate draws from these distributions for each of 400 sets of means and SDs. We truncate distributions so they are positively valued.[10] Figure 1 indicates whether we can reject the null hypothesis that the random draws' last digits were distributed uniformly: A gray circle means that we cannot rule out that last digits occurred with equal frequency, whereas a black X indicates otherwise. Blank space suggests that a distribution does not yield the relevant combination of mean and variance.

All distributions for which we simulated draws result in equal-frequency last digits for most moments, including the mixtures that Mebane (2006, 8–9) proposed as examples in which the second digit follows the Benford distribution. There are three situations in which we do not observe equifrequent last digits: First, this is true for distributions with very low SDs (less than 10, as a rule of thumb) because draws from such distributions cluster within a very narrow range of numbers. Second, it is also true for distributions that have a fixed upper bound (e.g., the same number of votes across polling stations) and draws that cluster at this upper bound (e.g., all or almost all votes are for one party).[11] Third, it is true for distributions with means that are low relative to their SDs because such distributions generate a large number of very small counts.

The empirical data discussed in this paper are not drawn from distributions of this kind: Vote returns we analyze vary substantially more than required, and they usually have means that are larger than the

---

[10]Means range from 1 to 200 and SDs from 1 to 100. They identify distributional moments prior to truncation. We simulate three runs of 1000 random draws from each distribution. For each draw, we compute the chi-square statistic given the null hypothesis of equifrequent last digits, and we conclude that last digits are not equifrequent if we can reject the null at the 95% level in all three runs.

[11]This case only arises in Mebane's mechanism A, where we see nonequifrequent last digits for example for a distribution with a mean of about 170 and an SD of about 75. This mechanism presumes that all precincts are exactly the same size, which is not the case in our data and probably in most election data. Even minor variations in precinct size (on the order of just tens of votes) will restore equifrequent last digits.
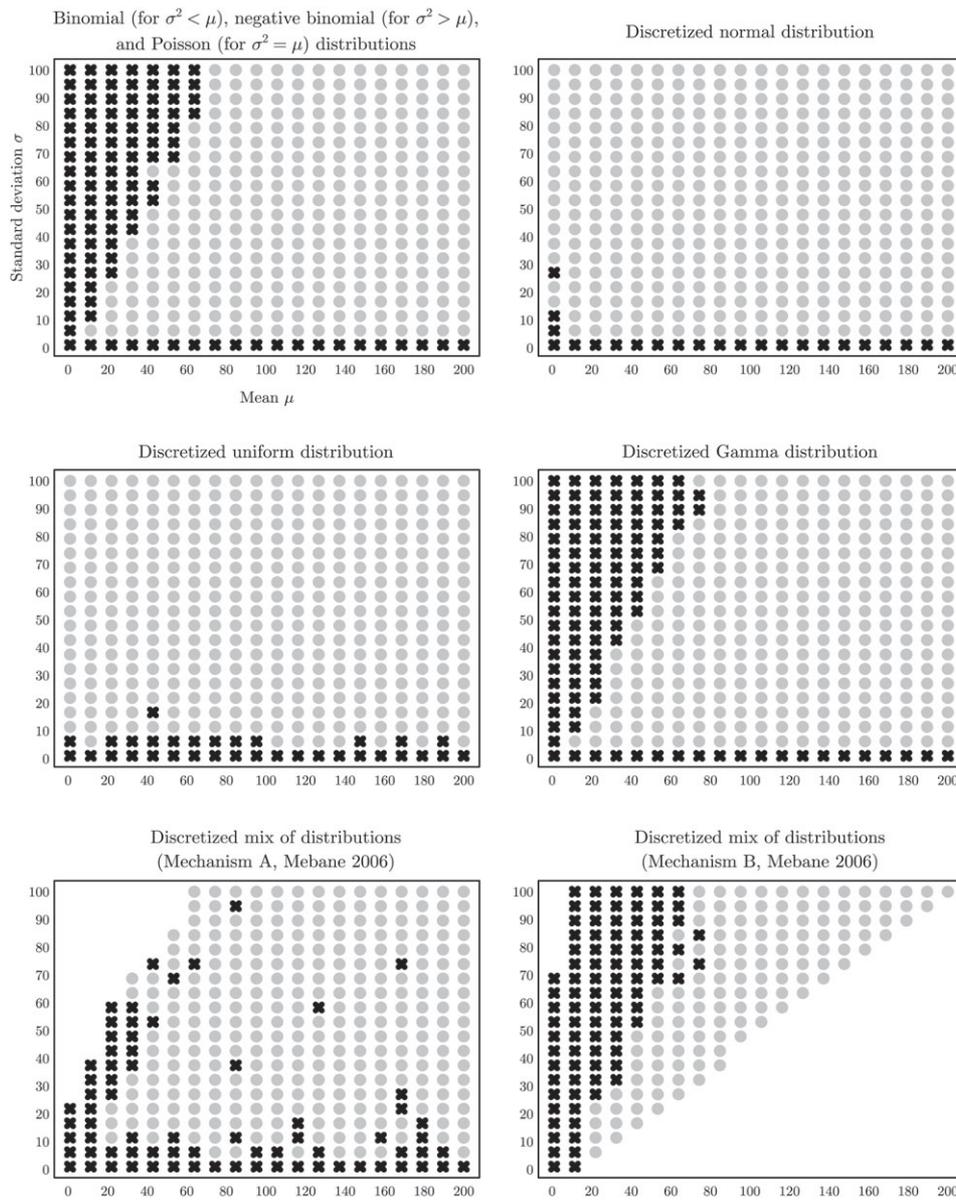
Fig. 1 Theoretical distributions and uniformly distributed last digits.

corresponding SDs (e.g., the average polling station vote count for Nigeria's incumbent party is about 270 in our data, with an SD of roughly 180).[12]

The simulations underline the fact that we are agnostic about the exact statistical process by which electoral returns are generated. One key advantage of our approach is that we need not be convinced that election results follow any particular distribution (or mix of distributions) because the theoretical result of uniformly distributed last digits holds across a wide range of data-generating processes.

---

[12]One exception are vote returns for the All Nigeria People's Party (ANPP), the main opposition party in the Nigerian election data we analyze, with a mean of 124 and an SD of 142. As a robustness check, we ran our analysis while excluding small ANPP returns (omitting single-digit draws in one specification and both single- and double-digit draws in another specification) and obtained substantively equivalent results. Results are available upon request.

2.3 *Deviations from Uniformly Distributed Last Digits*

The previous two sections indicate that we can generally expect election returns to yield uniformly distributed last digits, and we can informally summarize two scope conditions: First, we require that vote counts do not cluster within a very narrow range of numbers. For example, such clustering could occur in candidate returns if a large portion of polling stations serves about the same number of constituents at about the same level of turnout and support for the candidate. The simulations in the previous section suggest that even minor variation in election unit sizes or turnout rates can restore equifrequent last digits, so we suspect that this scope condition will rarely pose a serious constraint in practical applications.[13]

Second, we require that vote returns do not contain a large proportion of single- and double-digit counts. For example, the tests we propose should not be applied to the low vote counts of minor candidates or returns from very small polling stations.[14]

If these scope conditions are met, nonuniformly distributed last digits are unlikely to have been generated by a fair and proper election. But this in itself does not necessarily imply that they are the result of fraudulently manipulated numbers. Two other statements have to be correct in order for fraud to be probable: First, it has to be true that we can expect fraudulently produced numbers to yield nonuniformly distributed last digits. If rigged elections are as likely as fair elections to result in equifrequent last digits, then deviations from this expectation are not informative about the probability that election returns were manipulated. Second, it has to be the case that we can exclude other nonfraudulent irregularities (such as rounding) as alternative causes of nonuniform last digits (Preece 1981).

We provide an extensive discussion of the latter concern when we turn to the empirical analysis below. Here, we address the former issue and provide evidence that fraudulent numbers sometimes do go hand-in-hand with nonequifrequent last digits. We highlight a set of findings from the Office of Research Integrity (ORI), which oversees, supports, and sets policies for investigations of alleged misconduct in health research. The ORI forms part of the Department of Health and Human Services and directs integrity activities for a substantial share of federal research grants, including research sponsored by the National Institutes of Health and the Centers for Disease Control and Prevention.

In 2002, ORI staff published summaries of three cases in which it alleged that researchers had fabricated numeric data or results (Mosimann et al. 2002): a case of falsified hand written scintillation counts, which had been recorded even though measurements were known not to have taken place; a published set of means and SDs for endotoxin levels in cell cultures, which was called into question by a colleague and for which supporting raw data were lacking; and a set of measurements of weight-adjusted blood flow in rats, which appeared suspicious when blank raw data sheets and undissected carcasses were discovered.[15]

Two aspects of these cases are of particular interest in the context of this paper: First, in each case, trailing digits in the allegedly made-up data were shown to be nonuniformly distributed, in contrast to digits in a set of data not under suspicion from the same case, and the ORI interpreted these findings as a strong indication of misconduct. Second, in each case, the researcher under investigation admitted wrongdoing.[16]

This means that in these studies digit manipulation is known to have occurred and to have resulted in nonuniformly distributed last digits, whereas last digits in comparable uncompromised data from the same studies were uniformly distributed. The ORI analysis provides us with a rare comparison of unsuspicious

---

[13]This scope condition does not constrain the empirical analysis presented in this paper. In the case of Nigeria, for example, polling station size, turnout, and support for different parties all vary tremendously. A single polling station can serve anywhere from just a handful to over 1300 registered (and 1100 actual) voters.

[14]Note that we do not require electoral results to be generated or analyzed in any particular order since we only analyze digit pairs fully contained within a particular count.

[15]The article includes one additional case, which we omit because the researcher under suspicion in that case did not admit wrongdoing. Mosimann et al. (2002) do not explicitly discuss the ORI's most notorious case, which also relied in part on digit-based inferences and in which the ORI's findings were overturned on appeal (Kevles 1998). Al-Marzouki et al. (2005), Smith and Godlee (2005), and White (2005) provide and discuss an example of a non-ORI investigation of alleged misconduct in medical research, an investigation that also relies in part on an analysis of last-digit frequencies.

[16]In the first case, the researcher under investigation "admitted to the falsification of the numbers" (Mosimann et al. 2002, 78); in the second case, the respondent "agreed that the data were 'flawed' and retracted the relevant articles" (83); and in the third case, the accused investigator "admitted to falsifying data" (89), then withdrew his admission, but "accepted the finding of scientific misconduct" when he was presented with the ORI's analysis (92).

numbers with numbers we know (and do not only suspect) to have been falsified, and this comparison suggests that it is reasonable to expect fraudulent last digits to be nonuniformly distributed.[17]

The cases of research misconduct reported above suggest nonuniform last-digit distributions as an indicator of fraud. We have also implied that the fabrication of numbers in medical research is in important ways similar to the fabrication of election returns. In particular, we contend that they are similar with respect to how individuals make up numbers that are entirely unimportant for what he or she is actually trying to accomplish by cheating. It may seem counterintuitive to focus only on digits that do not make a difference as to whether a paper gets published or who wins an election. After all, a fraudulent terminal digit in a vote count can by itself yield no more than nine additional votes. But the problem with focusing on leading digits that ''make a difference'' is that the leading digits that fraudsters strategically select and their distributions in the absence of fraud will often be context dependent. The distribution of first digits of coefficients is probably different from the distribution of first digits in election counts in an election, which may be different from the distribution in another election.

Last digits of a sufficiently long number, on the other hand, are the equivalent of inconsequential noise, whether the number is a coefficient or a vote count.[18] And an individual making up such a number has no incentive to generate anything other than a noisy (more precisely, uniformly distributed) last digit, regardless of context. In turn, the fact that individuals, who have admitted to the manipulation of research data, were unable to generate such last digits suggests that the same may be true for cheating individuals in other contexts. As long as election officials are making up vote returns wholesale (and without the aid of number-generating algorithms), it does not matter what strategy they use to fabricate the leading digits that will actually decide an election.[19] The tests proposed here can be applied regardless of the official's (possibly context-specific) leading-digit strategy.[20]

This does not yet answer why the individuals in the cases documented above performed so poorly at the production of uniformly distributed last digits. We argue in the next section that this reflects ubiquitous human biases in the fabrication of numbers and is not a phenomenon limited to instances of misconduct in medical research or any other specific application.

3   Psychological Biases in Number Generation

Although we would expect many distributions to produce equifrequent last digits, humans have difficulties reproducing such random patterns, even when they have incentives to do so. We find evidence for this claim in a host of studies in psychology and statistics, which address ways in which people perceive randomness and their ability to produce random sequences in experimental settings (see Nickerson 2002 for a review). These studies focus mostly on binomial trials, and usually ask subjects to produce sequences of heads and tails to simulate results from coin tosses, but a small set of experiments directly addresses the problem of generating sequences of decimal digits.[21]

In the first such experiment, Chapanis (1953) asked 13 subjects with varying levels of formal education in statistics and probability theory to write out sequences of at least 2520 single digits. Subjects were asked to write down the digits 0–9 in ''random order'' on sheets of paper, without interruption, at their own pace. Chapanis found that subjects displayed marked preferences for certain digits, although they disagreed in their preferences. Without exception, subjects avoided repetition in their sequences, rarely creating repetitive pairs or triplets (such as 111, 110, or 101). They also highly preferred decreasing serial sequences (e.g., 987), although the same was not true for increasing sequences (e.g., 123). Digit choices were autocorrelated to the fourth order.

---

[17]More precisely, the prior probability that fraudulent numbers entail nonuniformly distributed last digits is bounded away from zero and could be large, which is not the case for the corresponding prior probability for nonfraudulent numbers.

[18]Mosimann, Wiseman, and Edelman (1995, 33) refers to such last digits as ''error digits.''

[19]Note that we focus on identifying fabricated counts, not cases of ''switching a proportion of votes from one party to another'' for which Cantú and Saiegh (2011) propose their leading-digit analysis.

[20]To be clear, we do not argue that officials are trying to sway elections by manipulating only last digits. We are concerned with officials fabricating numbers in their entirety, so that the leading digits win or lose the election but the trailing digits exhibit patterns we can detect by way of our tests.

[21]See Yule (1927) for an early, nonexperimental discussion of humans' decimal digit preferences.

In a similar experiment, Rath (1966) asked 20 university students to each produce 2500 random digits by filling in 10 one-page grids of 250 rectangles each. The students were told that they could leave the experiment as soon as they filled in all the sheets. Rath found four clear patterns in subjects' digit choices. First, subjects preferred some numerals over others, and they significantly preferred small numbers (1, 2, and 3) over both larger numbers (5, 7, 8, and 9) and zero. Second, he found strong biases against repetitive pairs of digits. Third, subjects exhibited strong biases in favor of adjacent pairs of numbers (such as 12 and 23). Fourth, digit pairs of distant numerals (such as 38 and 42), that is, numerals that are neither identical nor neighboring, occurred significantly less frequently than expected.

Boland and Hutchinson (2000) echo these findings in an experiment with 458 university students, where each was asked to produce random sequences of 25 single digits. Again, subjects preferred small digits (particularly 1, 2, and 3) and avoided others (5, 6, 8, and 0). They also find a strong effect for the avoidance of repetition. A striking 70% of respondents failed to repeat a single digit in their 25-digit sequence, whereas this would be expected to occur only 8% of the time using a random number generator.

One could argue that perhaps the biases uncovered in these experiments are the result of research subjects writing numbers carelessly and expediently in order to leave the site of the experiment as quickly as possible, a process that is potentially quite different from the intentional falsification of election returns. But note that subjects in Boland and Hutchinson (2000) do not have an incentive to finish as quickly as possible since they are students participating in the experiment as part of a scheduled class session of fixed length, and results are similar to the other two experiments. Biases in number generation are also no less prevalent in experiments that provide different monetary payoffs for effective randomization (Budescu 1987; Mosimann, Wiseman, and Edelman 1995; Rapoport and Budescu 1997).

The experiments discussed so far instruct subjects to write sequences of single decimal digits, whereas election officials write sequences of numbers that usually consist of more than one digit. Two other sets of experiments by Mosimann, Wiseman, and Edelman (1995) and Diekmann (2007) assign tasks more like the one faced by cheating officials, with similar results. Diekmann (2007, 327) asked 37 students, mostly of sociology and economics, to fabricate between 10 and 100 four-digit regression coefficients that would support a particular claim. He finds that last-digit frequencies significantly deviate from expectation, which is not the case for a sample of regression coefficients drawn from articles published in the *American Journal of Sociology*.[22]

Mosimann, Wiseman, and Edelman (1995) asked a total of 46 subjects, about half of them students and half of them research scientists and administrative staff, to make up a total of 90 three-digit numbers (60 of them with the numerals 0–9 and 30 with the numerals 0–2). They were asked to imagine that they were "in charge of a state lottery and . . . required to keep a record of the numbers." When some numbers turn out to be missing in this scenario, "[y]ou decide to make up . . . random numbers to replace the . . . missing numbers" (37). Again, subjects generally preferred low numbers (1, 2, and 3). The number 0 was avoided for the leftmost and rightmost position but was one of the most frequently chosen digits for the middle place (together with 1 and 2). Although the authors did not analyze serial relationships between digits, they investigated and found no correlation between a subject's ability to produce random digit draws and his or her student status, educational attainment, level of scientific training, or completion of a college-level math or statistics course. Ostensible differences in sophistication did not change subjects' ability to randomly select digits, which suggests that human biases in the production of numbers do not just affect individuals in particular circumstances or with a certain background.[23]

In summary, we focus on four findings that the experimental literature suggests: Humans (1) do not select digits with equal frequency, (2) avoid repetition, (3) prefer serial sequences, and (4) select pairs of distant numerals relatively infrequently. The first two findings in particular are consistent with a larger theoretical and experimental literature on cognitive biases in probability perception, such as the "gambler's fallacy" (where people expect the second draw of a signal to be negatively correlated with the first)

---

[22]Diekmann checks for deviations from the Benford distribution, which approaches the uniform distribution for higher-order digits. The expected last-digit frequencies for four-digit Benford distributed numbers are nearly uniform and range from .09982 to .10018 (Diekmann 2007, table 1).

[23]Stress also does not appear to have a significant impact on individuals' ability to randomize (see Kuhl and Schönpflug 1974, who induce stress by exposing subjects to noise up to the level of a power mower).

and beliefs that very small samples resemble the parent populations from which they are drawn (Tversky and Kahneman 1972). Drawing on these studies, we test for the four patterns identified above.

4 An Empirical Assessment of Electoral Fraud

4.1 *An Empirical Baseline: Sweden's 2002 Parliamentary Elections*

We establish an empirical baseline, a null result against which our later analysis can be compared, using data from the 2002 parliamentary elections (Riksdagsval) in Sweden.[24] We analyze returns at the ward level (valdistrikt), where the 5976 wards in our data set are nested within 290 municipalities (kommun) and 21 counties (län). As far as we can tell, there were no suspicions of electoral fraud or return sheet manipulation raised with respect to this election, and our theoretical expectation is that last digits will be distributed uniformly. Figure 2 shows last-digit frequencies across all wards for four different data columns: votes for the Social Democratic Party (Socialdemokratiska Arbetareparti [SAP]), votes for the Moderate Party (Moderata Samlingspartiet [MSP]), the number of registered voters (antal röstberätti-gade), and total votes cast. Horizontal lines indicate the expected value of .1 as well as the lower and upper confidence bounds.[25] Our expectation is confirmed: Each numeral appears roughly as often as any other.
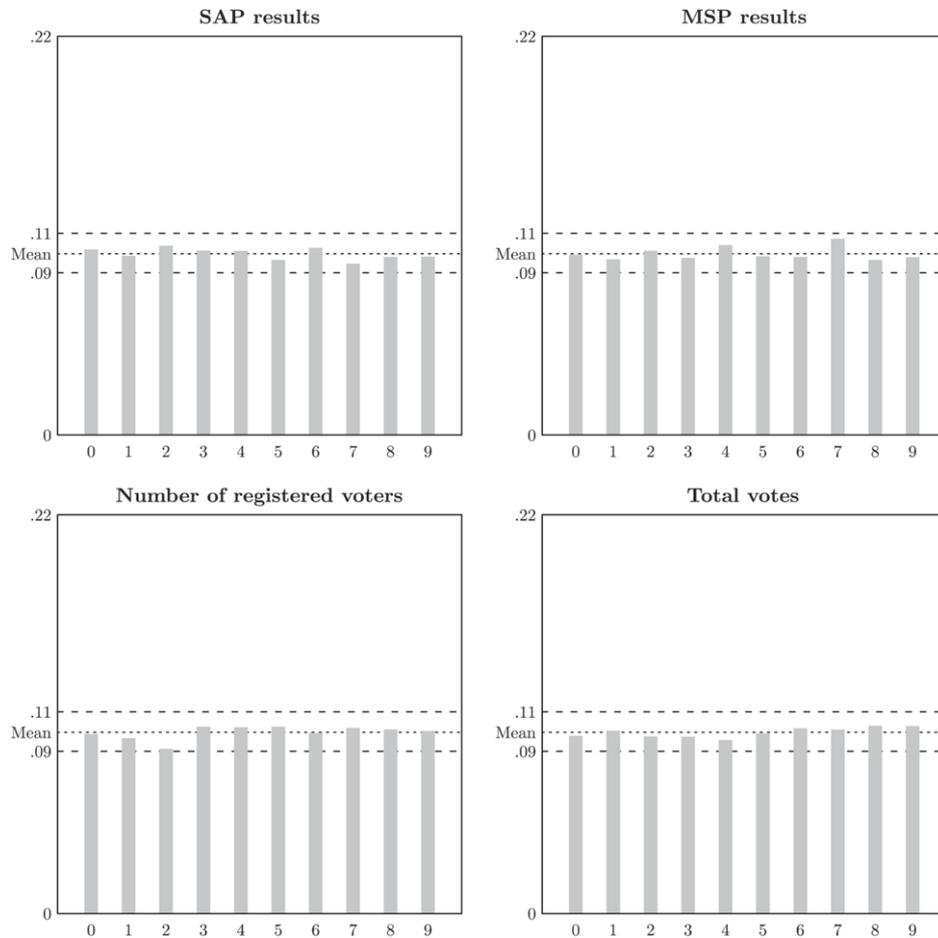

Fig. 2 Frequencies of last digits, Sweden 2002.

---

[25]Confidence intervals are constructed so that the probability that an untampered set of vote counts yields a digit frequency below or above the confidence bounds is .05.
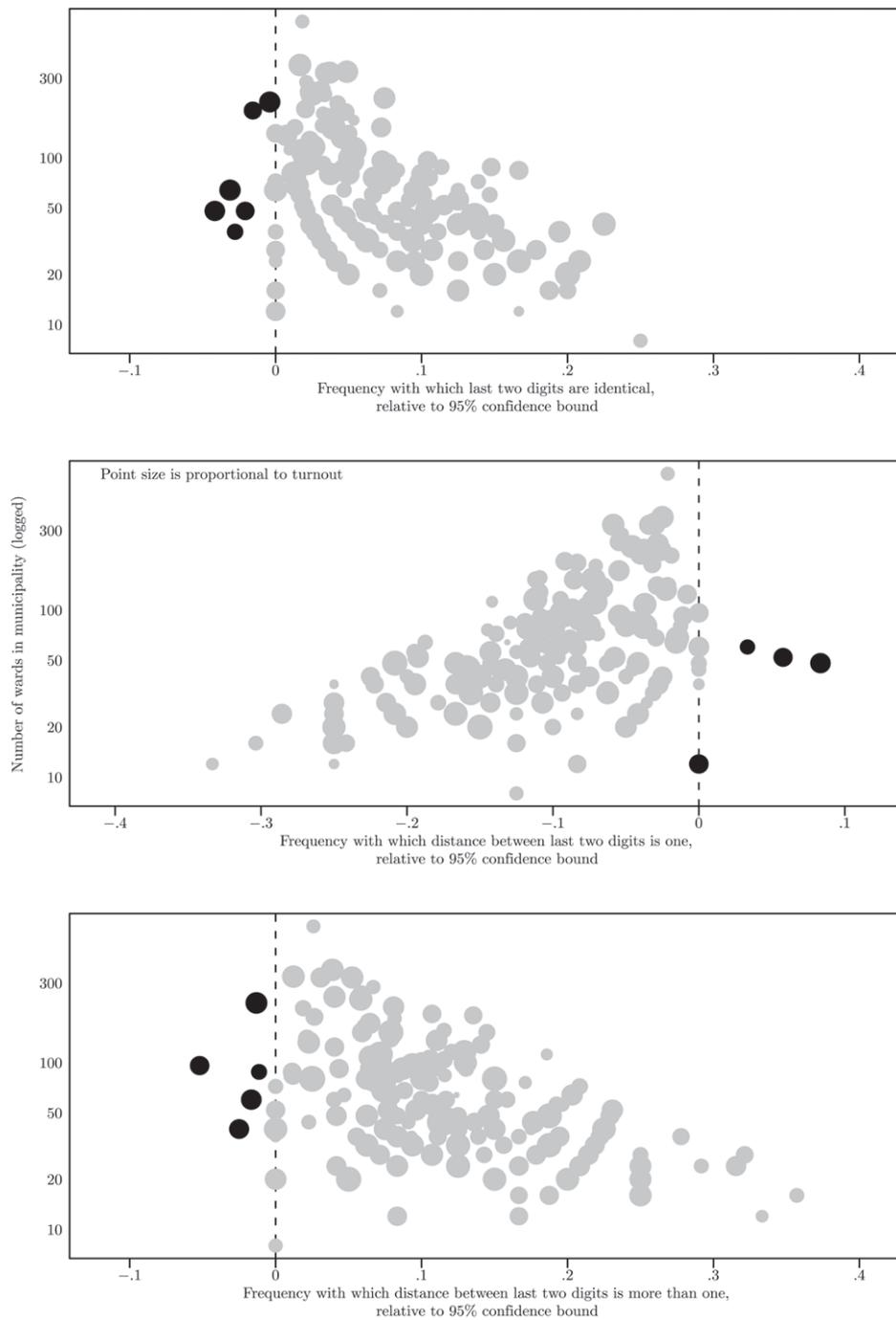
Fig. 3 Last and penultimate digits, Sweden 2002.

Similarly, absent fraud, we would not expect to observe a conspicuous lack of digit repetition, an improbable glut of adjacent digits, or a dearth of distant numerals in pairs of last and second-to-last digits. Figure 3 provides evidence to that effect. Each point represents one of the municipalities, which are sorted by size (i.e., ward count) along the vertical axis. Point size is proportional to turnout, which we measure as the share of registered voters who cast a ballot. We consider here pairs of last and penultimate digits found in the columns for the SAP vote count, the MSP vote count, the number of registered voters, and the total number of votes within each municipality.

The horizontal axis gives the quantity of interest. For the first graph, it denotes the extent to which the last and penultimate digits within a given municipality are the same, relative to the lower confidence bound. Municipalities marked in black to the left of the dashed line at 0 have suspiciously few repetitions. The second graph shows the extent to which digits are adjacent in a given municipality, relative to the upper confidence bound. Points above 0 have worryingly many pairs of adjacent digits. The third graph displays the degree to which we observe pairs of nonneighboring and nonrepeating digits, relative to the lower confidence bound. The black dots indicate municipalities with suspiciously few pairs of distant digits.

There are a small number of municipalities that are seemingly suspicious, but this is the result of the fact that we plot unadjusted 95% confidence bounds for a test of many hypotheses—one for each municipality. Since we plotted just over 200 municipalities (in order to facilitate comparison with our analysis of Nigeria's 2003 election), it is not surprising that a small number of them will lie beyond the 95% confidence interval purely by chance. Again, we cannot reject the null hypothesis of a "clean" election.

4.2 *A Case of Alleged Fraud: Nigeria's 2003 Presidential Elections*

We now use our digit-based test to examine electoral returns from Nigeria. We analyze data at the polling station level for Plateau state, which is located in the "middle belt" region of the country. We personally retrieved these data, and it remains, to our knowledge, the first time that postcolonial election data at this level of aggregation has been available outside of Nigeria. Here, we analyze presidential election returns.

All results were entered from original, handwritten electoral ward report sheets used by local authorities, and we focus on the ward as our target of analysis.[26] For the vast majority of sheets, the handwriting suggests that only a single official entered the results, so these returns are well suited to test whether we can discern fraudulent sheets by leveraging individuals' psychological biases. We analyze results from all of Plateau state's 204 wards, where each ward contains an average of about 13 polling stations. Ward size varies significantly: Wards in the bottom decile contain no more than five polling stations each, whereas those in the top decile are each comprised of more than 22 polling stations. In turn, the power of our analysis varies substantially across wards.

It is no secret that electoral fraud occurred during the 2003 elections in Nigeria and that some ward results were manipulated wholesale. At the time of the 2007 elections, *The Economist* reported that in Anambra state, "[b]arely any polling-stations got a results-sheet . . . ; presumably these were being filled in elsewhere" (*The Economist* 2007). Similarly, "the head of Nigeria's Catholic Bishops Conference . . . cited massive fraud and disorganization, including result sheets being passed around to politicians who simply filled in numbers at will while bribed returning electoral officers looked away" (Agence France Presse 2007). We believe the fact that it is extremely likely that fraud occurred in Nigeria makes the country an especially useful case for validating our digit-based approach. We have used data from Sweden to illustrate that our digit-based test does not ring alarm bells in an apparently nonfraudulent environment. Data from Nigeria will help us establish that our test is in fact sensitive to real manipulation of election results.

In the Nigerian context, covariate-based tests for electoral fraud face serious challenges. Very little data are available at the substate level and none at the ward level. Census data are notoriously unreliable—the most recent census prior to 2006 took place in 1991 and was marred by blatant fraud and large-scale violence—and few would argue that it accurately reflects the facts on the ground. In this environment, monitors and outside observers need other tools to assess patterns of fraud.

A digit-based approach helps monitors and observers in two other ways as well. First, some areas may be unsafe for election monitors to visit, so tools for detecting fraud other than direct observation are needed. Second, if observers help curtail fraud techniques such as intimidation and ballot box stuffing at the polling station level, they will also need tools to validate that elites have not simply shifted to manipulating return sheets.

---

[26]An example of such a ward-level return sheet, which highlights the data columns we will analyze below, is available in the supplementary materials on the *Political Analysis* Web site.

Why does it make sense to look at ward-level manipulation in Plateau state, in particular? In a paper on the 2003 elections, Darren Kew groups Nigeria's states into three broad categories of fraudulence (Kew 2004):

- In a handful of states, including Lagos state, minor violations occurred across polling stations, but "most did not detract from holding serious elections, with results at the polling stations and wards that were generally accurate" (157).[27]

- Another dozen states saw "fairly credible elections, but with some instances of rigging in rural locations. Moreover, their results grew increasingly questionable as they moved through the LGA [local government area] and state collation centers" (157). In Adamawa, for example, People's Democratic Party (PDP) operatives appear to have altered returns at collation centers, and Kew suggests that the PDP may have "originally planned to fudge the numbers" (159). He includes Plateau state in this middle category.[28]

- In another third of Nigeria's states, leaders of the incumbent PDP "did not even attempt to erect a facade of validity" (161). Kew writes that in Rivers state, for example, "all the international monitors stationed across the state, myself included, directly observed widespread ballot-stuffing for the PDP" (161). Other types of fraud documented in these worst-offender states included rampant underage voting and outright physical intimidation of voters and monitors at polling booths.[29]

The second group of states is precisely the context where our application—looking for manipulation at local collation centers rather than at the polling stations themselves—is most useful. The small number of states in the first category may have had little fraud at all, whereas in states in the third category, manipulation of digits at local collation centers is likely to be overshadowed by blatant fraud at polling stations.

4.2.1 Last-digit frequencies

We first examine the frequencies with which different numerals appear in the last digit. Figure 4 provides digit frequencies across all wards for four different return sheet columns: votes received by the incumbent PDP; votes received by the main opposition party, the ANPP; the number of registered voters; and total vote counts. The contrast with Fig. 2 is striking. For all four columns, we observe significant deviations from the uniform distribution (which are marked in black), in particular for the numeral 0. This strongly suggests that electoral returns were indeed manipulated.[30]

Another way to test whether last digits depart significantly from the uniform distribution is to measure the extent to which digit frequencies vary. We compute the chi-square statistic and corresponding p-value for each data column: The probability that a "clean" election would produce at least as much variation in digit frequencies as we observe is about .01 for PDP returns and less than .0001 for ANPP returns, the number of registered voters, and total vote counts.[31]

The fact that we find deviations from uniformly distributed last digits across all wards suggests that manipulation of return sheets was widespread. If numbers had been manufactured only in some isolated wards, this should not lead to such a strong result in the aggregate data presented in Fig. 4.

---

[27]Support for the PDP in Lagos reached 69.35%, which is low in comparison to other nearby states. In Ogun, the official result records a PDP vote share of 99.9%; in Ondo, 94.6%; in Osun, 95.2%; and in Oyo, 93.9% (Kew 2004, 153).

[28]In some instances, the opposition ANPP "seems to have taken a similar approach," for example, in Kebbi and Yobe states (Kew 2004, 160). The PDP received 67.3% of the official vote in Plateau, compared to 36.7% in Bauchi to the north, 53.5% in Kaduna to the west, 64.6% Nasarawa to the south, and 76.7% in Taraba to the east (152–3).

[29]This group of states includes in particular the coastal states in the southeast, all of which reported exceptionally high returns for the PDP (93.9% in Delta, 96.0% in Bayelsa, 92.8% in Rivers, 83.9% in Akwa Ibom, and 97.9% in Cross River). Across the country, PDP candidate Obasanjo received 61.9% of the official vote count.

[30]One question this raises is why officials did not just change vote returns' leading digits since those are the digits that decide elections. First, this alternative process cannot explain the departures from expectations that we observe for last digits. Although it is plausible that some election officials will cheat in this way, this is not, or not only, what happened in the case of Nigeria. Second, one reason why officials may fabricate numbers in their entirety is because they block a truthful vote count in order to prevent a record of the correct tally or because they submit results before a count has been conducted.

[31]Even if we exclude counts that end in zero, last digits vary substantially more than in the case of Sweden. A chi-square test of the frequencies of nonzero last digits across all four columns yields a p-value of about .11, short of statistical significance. The corresponding p-value for the data from Sweden is .97.
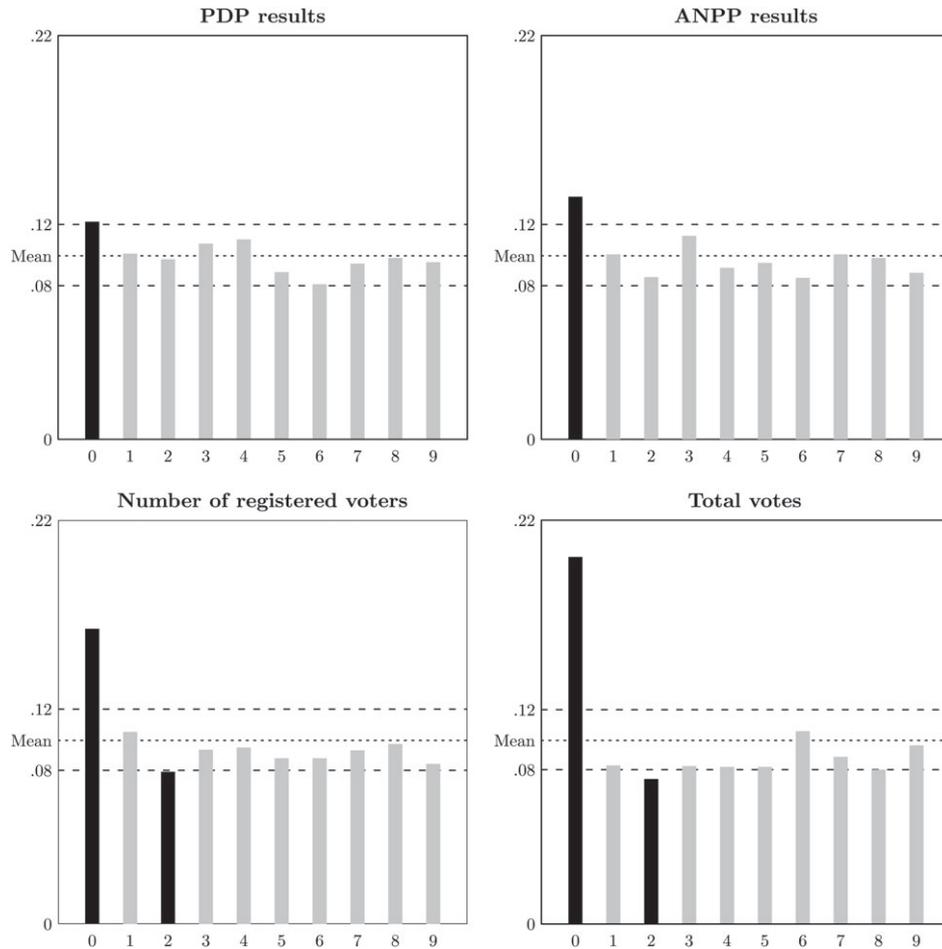
Fig. 4 Frequencies of last digits, Nigeria 2003.

We obtain significant results not only across all wards but also within a significant number of them. The right panel of Fig. 5 shows a histogram of *p*-values from ward-by-ward chi-square tests of the null hypothesis of equally frequent last-digit numerals. The left panel shows the equivalent histogram for a set of municipality-level tests of the data from Sweden, for comparison. The histograms group *p*-values in five bins: values at or below .05 (highlighted in black), .1 (in dark gray), .2 (in light gray), .5, and 1. In the absence of manipulation, we should expect *p*-values to be approximately uniformly distributed. For example, we would expect the proportion of cases with *p*-values at or below .05 to be about .05, which is what we see in the case of Sweden. In the case of Nigeria, however, many more ward-level tests generate significant results than would occur by chance. There are almost three times as many *p*-values at or below .05 than we would expect to see in a fair election.[32]

Figure 6 again shows last-digit frequencies but across all four columns of interest and for a random sample of about a third of the total number of wards.[33] Frequencies marked with a black rather than gray line exceed the dashed 95% confidence bound.

---

[32]Given the small size of some wards and the correspondingly low power of the chi-square test, this is a conservative statement about the prevalence of manipulation in the Nigeria data. For some wards, large *p*-values could be the result of their small size rather than the absence of suspicious digit patterns.

[33]Wards were sampled within four strata defined by the presence or absence of addition errors and of a significant result on our last-digit test, where the number of wards sampled from each stratum is proportional to stratum size. This ensures that the sample shown in Fig. 6 accurately reflects properties of the complete data.
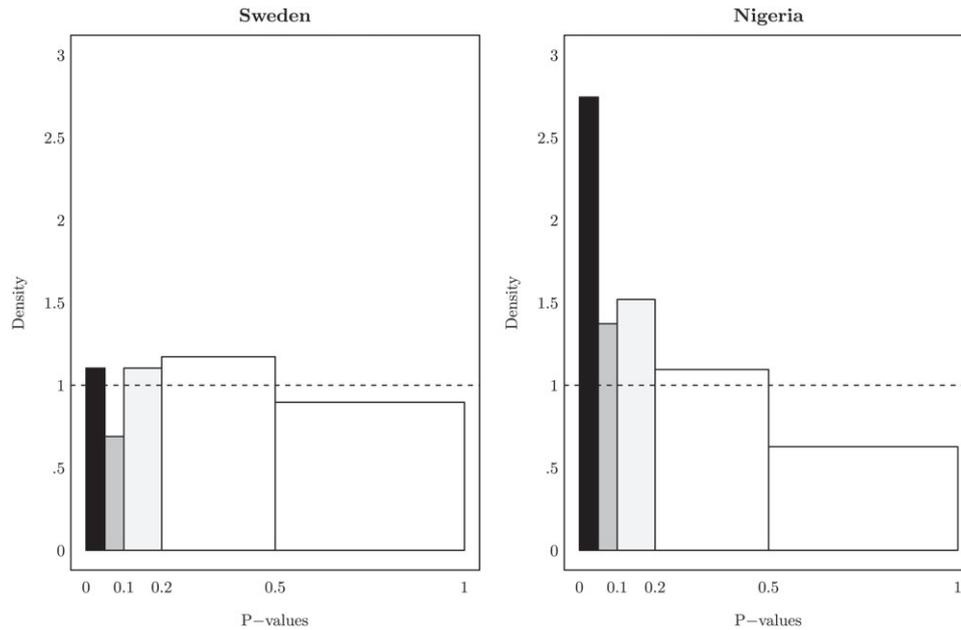
Fig. 5 *p*-values from ward-level chi-square tests.

Figure 7 shows a measure of the frequency of the most common last digit only, but it does so for all wards. Each point represents a ward, its horizontal position indicates its most frequent last digit, its vertical position shows the extent to which the last digit's frequency falls above or below the 95% confidence bound, and the point size is proportional to ward size. As before, we find suspiciously nonuniform distributions of last digits in a substantial number of wards.

This figure also provides information on ward turnout, by way of point shades. The graphs illustrate that there is no significant correlation between turnout and suspicious digit distributions, and logit regressions of indicators of potential manipulation on turnout produce statistically insignificant coefficients. Turnout is dubiously high in a number of areas, with about one in seven wards reporting turnout among registered voters to be higher than 95%, but these areas are generally not the same as the areas in which we find suspicious digit distributions.

We think there are at least two reasons for this disconnect. First, in the absence of information on demographics, we compute turnout as the ratio of total votes cast to registered voters. In and of itself, this procedure leaves us with an inflated estimate of turnout (the median ward turnout is about 82%), which limits the extent to which high turnout can usefully serve as an indicator of election fraud. This problem is aggravated by the fact that the registration process in Nigeria prior to the 2003 elections "was marred by a number of severe irregularities" (European Union Election Observation Mission 2003, 21), which may have prevented some 30–40% of eligible voters from registering (although Nigeria's Independent National Election Commission maintained that the register accounted for virtually all adult Nigerians). With registration figures artificially depressed, particularly among the less persevering, who would presumably have been the least likely to turn out to vote if they had been registered, our estimated turnout figures are inflated further.

Second, the lack of correlation between dubiously high turnout figures and our digit-based measure arguably points to the fact that these indicators are sensitive to different types of fraud. A measure focusing on last digits captures manipulation of return sheets by the individual writing in numbers. High turnout figures could be the result of underage or forced voting, ballot box stuffing, or cheating on report sheets. We isolate a specific type of fraud, which could potentially be traced to named individuals, but are not surprised to find that other mechanisms of fraud could play an important role in areas for which we cannot reject the null of a "clean" election.
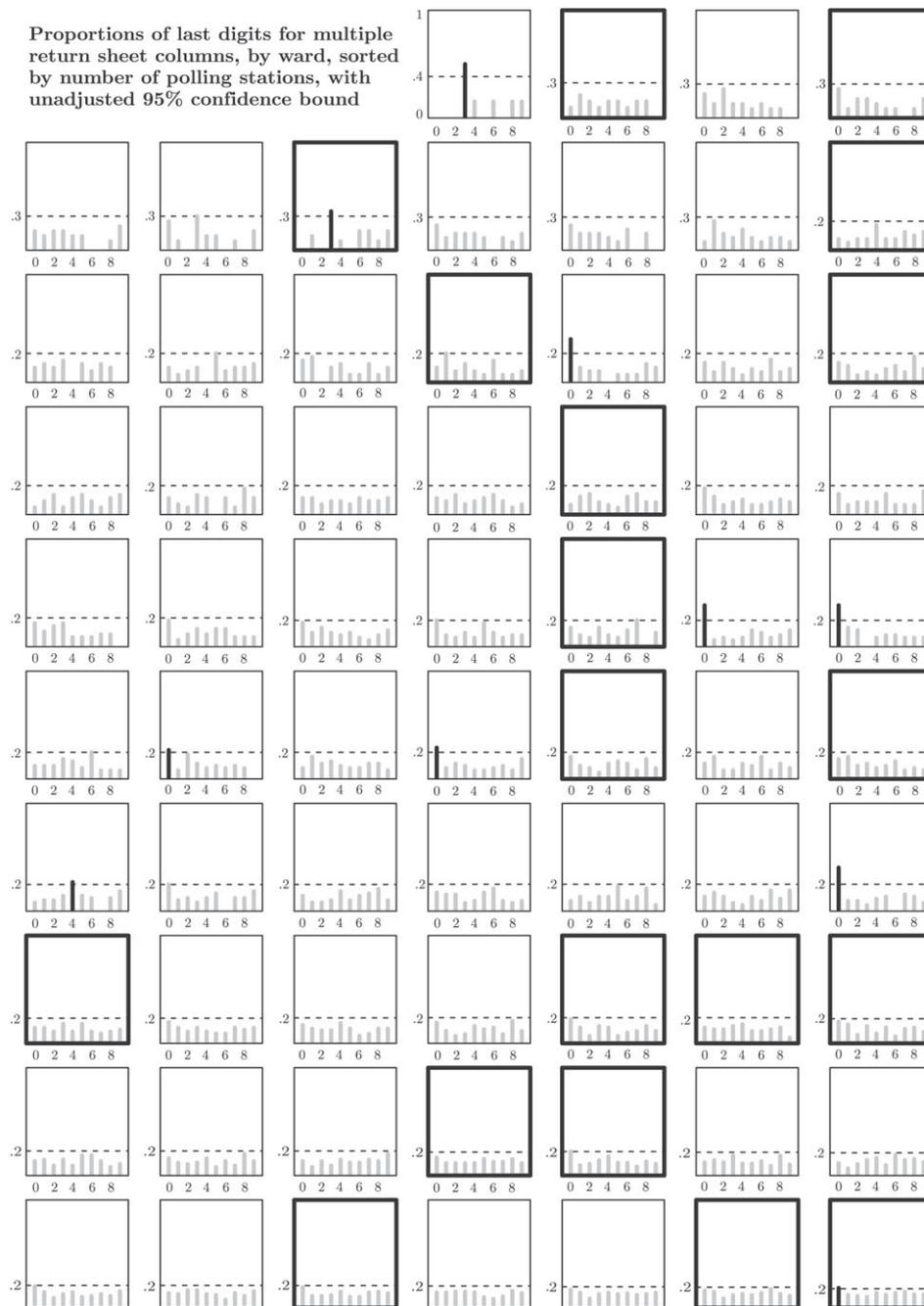
Fig. 6   Digit frequencies, wards with addition errors in bold.

4.2.2   Repetition, adjacency, and distance between numerals

We now assess the extent to which digit pairs exhibit repetition, adjacency, and distance of numerals across wards. Figure 8 uses data for Nigeria but is equivalent in design to Fig. 3. Against expectations from the psychology literature, we find no evidence of return sheets with too few digit repetitions and only weak evidence of an overabundance of adjacent digits. (Recall that we can expect a small number of wards to
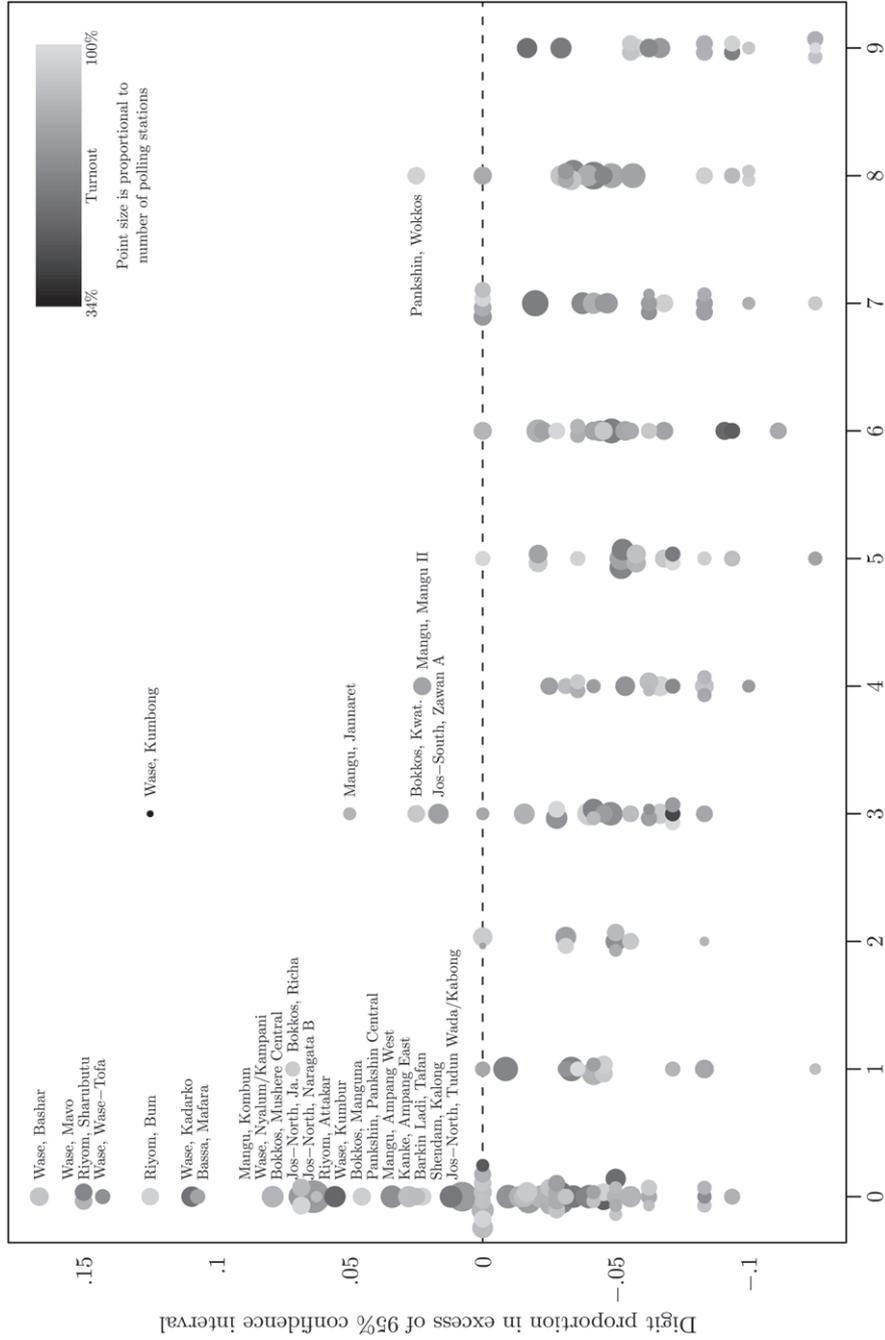
Most frequent last digit on ward return sheet

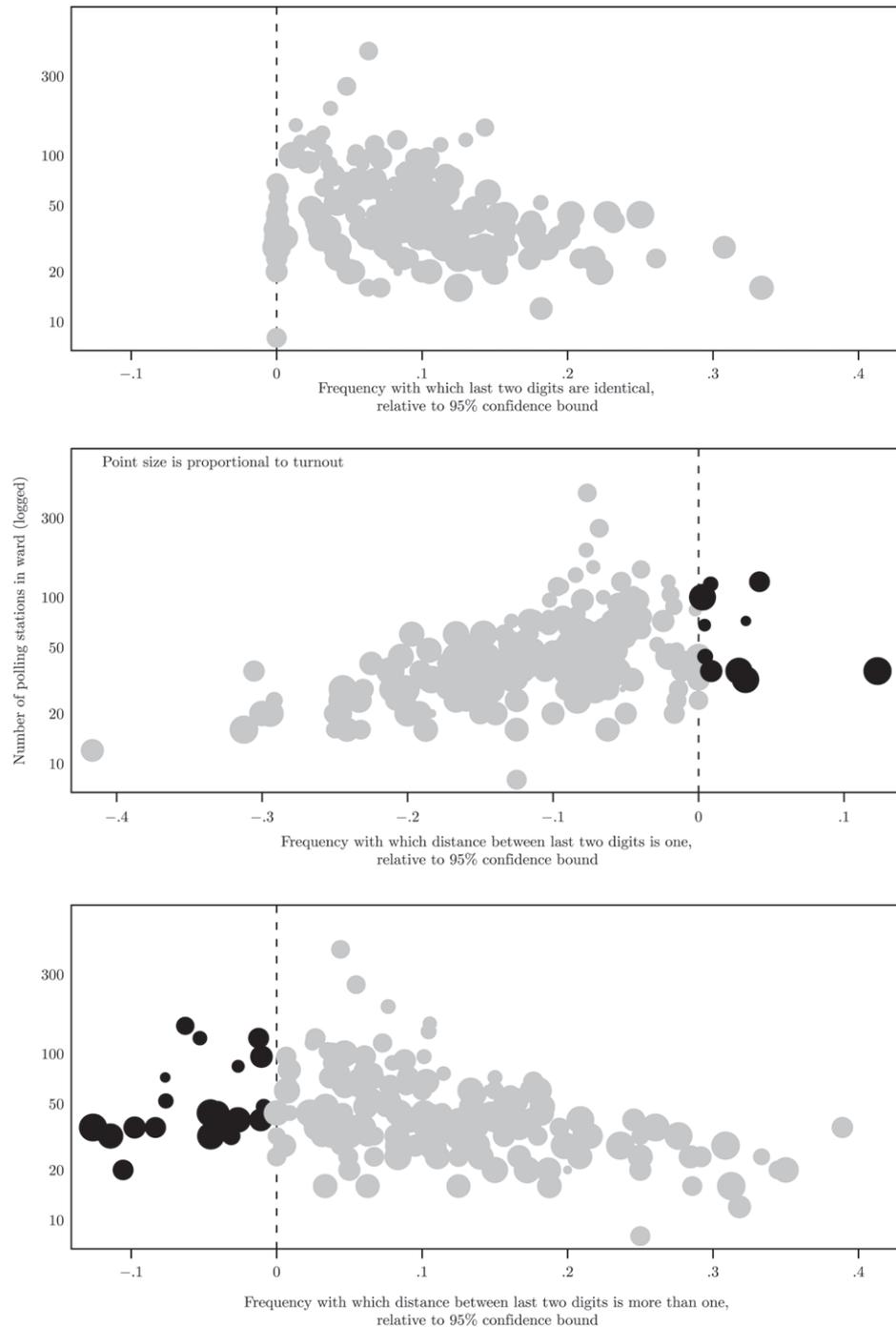**Fig. 7** Excess share of most common last digit.

Fig. 8 Last and penultimate digits, Nigeria 2003.

exceed the confidence bound purely by chance.) If anything, there appear to be too many instances of repeated numerals. We do, however, find that in a large number of wards, digit pairs do not often enough bridge a distance of more than one. When we look across several columns for each ward return sheet (PDP votes, ANPP votes, number of registered voters, and total vote count), we can identify 20 wards or almost one in ten in which pairs of distant digits occur suspiciously infrequently.

How can we reconcile this strong finding on missing pairs of distant numerals with the weak finding on pairs of adjacent numerals and the fact that digit repetition is at least as common as it should be? We argue

that this is largely a reflection of the statistical power associated with each measure. Even for relatively small wards, the expected number of pairs of distant numerals is fairly large, at least in comparison to the expected number of pairs of repeated or adjacent digits, and so the 95% confidence bound is relatively unforgiving. Small wards need to exhibit adjacency to a very substantial degree in order for us to be confident enough to reject the null hypothesis of a fair election, whereas a less extreme insufficiency of nonneighboring digits could push the ward beyond the 95% confidence bound.

Finally, our results are consistent with reports by international election monitors, who generally concluded that electoral fraud affected rural areas most severely. Most observers noted that flagrant violations of rules at the polling stations were much more likely to occur in rural as opposed to urban areas because international election observers were far more prevalent in cities (Kew 1999). We code wards as urban if they lie in a local government area that is synonymous with a major town as identified by the 2003 Nigeria Congress Online,[34] and a logit regression of an indicator of potential manipulation (which equals 1 for wards in which either last-digit frequencies or the frequencies of distances between last and penultimate digits deviate significantly at the 95% level from expectation in a chi-square test) on this indicator for urban areas produces a negative and significant coefficient ($-.97$, with SE .39 and a p-value of .01).

Although we find dubious digit sequences not only in rural but also in many urban areas (note that in Fig. 7 a number of questionable ward sheets come from Jos-North, a densely populated urban local government area), our analysis suggests that officials in rural areas apparently complemented visible violations of fair election procedures such as intimidation at the ballot box with vote return manipulation behind closed doors.[35]

4.2.3 Fraud or laziness?

In summary, last-digit frequencies that deviate from the uniform distribution remain our strongest evidence for the manipulation of vote counts in Nigeria. We observe such deviations in a substantial number of wards, many more than we would expect even when testing such a large number of hypotheses. In suspicious wards, zeros in particular are overabundant.

We argue that this pattern suggests fraud, but one could argue that many zeros simply indicate benign laziness as election officials are rounding to the nearest multiple of ten to facilitate their election day accounting.[36] One piece of evidence lends some support to this alternative hypothesis: Figure 6 sorts wards from smallest (fewest polling stations) to largest (most polling stations), and it appears that officials in larger wards are more likely to fill a return sheet with zeros. A logit regression of a ward-level indicator for a suspicious last digit (which equals 1 for wards in which at least one digit proportion deviates significantly at the 95% level from its expectation) on ward size yields a statistically significant and positive coefficient (.04, with SE .02 and a p-value of .03).

There are three reasons, however, why we are not convinced by this argument. First, we are more likely to reject the null of no manipulation in large wards for the simple reason that they are large, whether or not ward size is a predictor of digit modification. The more polling stations a ward contains, the tighter the 95% confidence bound around the uniform distribution of last digits, and the fact that suspicious digit distributions occur more frequently in large wards could be spurious to the fact that we face greater uncertainty in evaluating small wards. (Note that zero is the most common numeral for a number of smaller wards, even if its frequency does not surpass the relevant 95% confidence bound.) In fact, a logit regression

---

[34]Those towns are Jos, Bukuru, Barkin Ladi, Pankshin, Shendam, Langtang, and Vom. Nigeria Congress Online was a Web project sponsored by the Nigerian parliament and funded by the Department for International Development and the British Council. It included extensive information about Nigerian politics such as administrative structures, aggregate election results, members of the legislature, parliamentary committees, and Nigeria's constitutions. It is not available online but is available from the authors upon request.

[35]We find it plausible to think that in general different types of fraud can serve as substitutes for one another, even if this may not be the case here. See, for example, Ichino and Schündeln (2011) on the potential displacement of irregularities. One reason why officials may want to deploy several techniques of fraud simultaneously, complementing, for example, digit manipulation with ballot box stuffing, is to maximize their chances of obtaining a particular outcome, in particular if the marginal cost of engaging in an additional type of fraud is low. Another reason could be that some types of fraudulent behavior serve a purpose other than simply winning the election at hand, such as sending a signal of strength to and disabling an effective opposition.

[36]We focus here on rounding, but the same logic and arguments apply with respect to truncation and other appropriate floor and ceiling functions.

of an indicator that identifies wards for which a chi-square test of equifrequent last digits returns a significant result at the 95% level on ward size produces a statistically insignificant coefficient (.02, with SE .02). That is, once we use a more powerful test that compensates for the fact that suspicious digit frequencies can be more readily identified in wards with many polling stations, suspicious returns appear to be just about as likely in small as in large wards. This goes against the notion that rounding was behind the excess zeros in our data because in that case we would expect more digit manipulation in large wards with relatively more numbers that need to be processed.

Second, recall from Fig. 4 that we observe an excess of zeros in total vote counts in particular, but there are only a few polling stations (about 9%) in which all vote columns end in zero. (There are 19 such polling stations, and in all of them only votes for the two main parties ANPP and PDP were recorded.) But this implies that all other excess zeros, if they are the result of officials innocently rounding total vote counts, should go together with addition errors. Rounding of total votes has an easily observable direct implication: Party votes will not always add up to the exact figure reported as the total vote count. If some officials are simply taking a benign shortcut when adding party votes by choosing to round those counts to a multiple of ten, then the total vote counts reported by these officials will relatively frequently be different from what we get when we add party votes without rounding. If we do not find this to be the case in our data, we can rule out rounding as the principal source of excess zeros in total vote counts.

The evidence suggests the latter: Fig. 6 highlights wards in which we observe addition errors particularly often, that is, the reported total vote count differs from the accurately tabulated total vote count for more than 20% of polling stations, and these are not generally the same wards as the ones in which we see suspicious digit distributions. In fact, a logit estimate of this relationship gives a negative but insignificant coefficient (–0.66, with SE 0.52).[37] If anything, it seems, officials manipulating return sheets try to make sure that the numbers add up, which ironically makes their return sheets all the more suspicious given the otherwise high error rate.[38]

If rounding does not explain the abundance of zeros in vote counts, what does? One possible explanation is that the strong result for the total vote column in particular shows that officials combine fraud with expedience. As an official works to produce a ward sheet that sums correctly across columns, the official may fix a round figure for the total vote count and then work backwards to fill in the party vote counts that sum to this figure (perhaps working in this fashion precisely because entering round figures for party vote counts would appear relatively more suspicious). Total vote counts that end in zero are then not the result of benign rounding but lazy cheating.[39]

Third, we show in the following section that empirical evidence from cases other than Nigeria suggests that an abundance of zeros may be associated with fraud. We highlight data from Nigeria because it allows us to see whether our proposed test correctly suggests fraud when fraud has indisputably occurred. But it is perhaps problematic in that we do not have data for a local nonfraudulent comparison case. One could object that an abundance of zeros is perhaps the result of a local, nonfraudulent process that would also generate zeros in a fair Nigerian election, even if we do not find evidence of any such process in the case of Sweden. The next section focuses on the case of Senegal to counter this objection.

4.3 *A Within-Country Comparison: Senegal's 2000 and 2007 Presidential Elections*

We turn to an analysis of polling station data from the 2000 and 2007 Senegalese presidential elections.[40] Like Nigeria, Senegal is a sub-Saharan West African country at a low level of economic and human

---

[37]One could be concerned that this reflects a situation in which officials round only some party vote columns (as noted above, sheets with round vote counts for all parties are relatively rare), making it easier for them to correctly add up total vote counts, which they do not round. This could result in wards with suspicious digit patterns in party vote counts and few addition errors despite rounding. However, this is inconsistent with the fact that the relationship reported here is the strongest for wards with suspicious digit patterns in total votes. A logit estimation linking an indicator for suspicious total vote counts with a high incidence of addition errors returns a coefficient of –1.47 (SE 0.76), close to statistical significance at the 95% level.

[38]A reviewer points out that cheating officials could also have been instructed or selected for their ability to reconcile return sheets.

[39]Note that zero is not usually a preferred digit in the psychological experiments discussed in Section 3. But it seems plausible that it is a preferred digit in this context precisely because numbers ending in zero are easier to add up. As cheating officials try to make fake return sheets look proper, they ensure that numbers sum correctly and hence write in numbers that are easier to sum. There is also, in any case, disagreement across the small number of studies that have been conducted about which digits are preferred by subjects.

[40]This datum was obtained, with perseverance, by Jeffrey Conroy-Krutz, who generously agreed to share it with us.
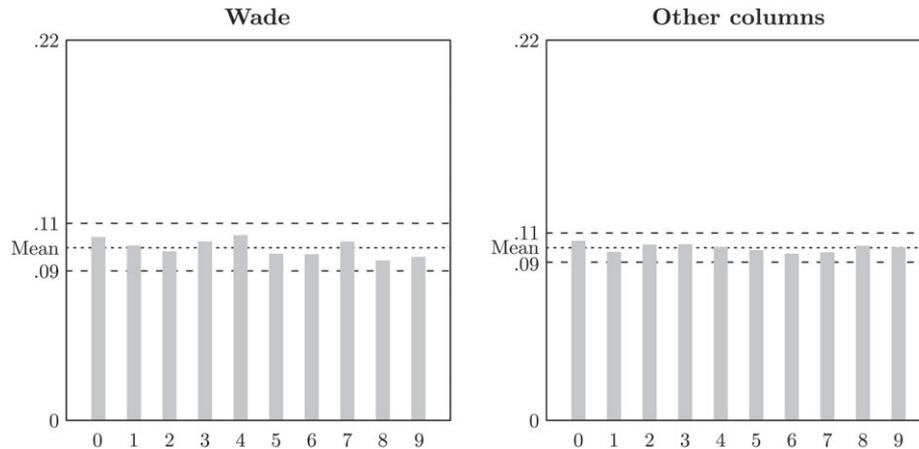
Fig. 9   Frequencies of last digits, Senegal 2000.

development: Senegal's 2007 per capita gross domestic product stood at U.S. $953, compared to Nigeria's U.S. $1123 (World Bank 2011), and Senegal was ranked 156th compared to Nigeria at 158th on the 2007 Human Development Index (United Nations Development Programme 2007).

Press and observer reports are not as clear-cut on the extent to which fraud tarnished the Senegalese elections as they are in the case of Nigeria. But news coverage in 2000 and 2007 suggests that the latter election was possibly tainted by manipulation and fraud, whereas the former is generally believed to have been clean. The election in 2000 resulted in a peaceful handover of the presidency from Abdou Diouf to Abdoulaye Wade, and then South African President Thabo Mbeki "congratulated the Senegalese people for their 'outstanding' conduct during the election process. . . . 'It is clear that the history of democracy is firmly established in Senegal and the whole election process could serve as an example for many countries in the world'" (Agence France Presse 2000). The 2007 reelection of Wade on the other hand generated considerable controversy, which ultimately culminated in a boycott of the parliamentary elections by most of the opposition less than four months after the presidential ballot.[41]

This means that in the case of Senegal we can compare a tentatively fraudulent election with one that was probably clean, which can strengthen our argument with respect to the data from Nigeria in two ways. First, if we find that suspicious digit patterns from the controversial 2007 election resemble those from Nigeria, where we observe excessive zeros, it becomes more likely that election fraud can indeed generate such patterns. Second, if we find that the relatively clean 2000 election fails to produce these patterns, it becomes less likely that some sort of nonfraudulent process could explain our results.

Figures 9 and 10 show last-digit frequencies for the two Senegalese elections. We show frequencies for the winner in both elections, Abdoulaye Wade, in the left panel. The right panel summarizes the remaining return columns with vote counts that were sufficiently large and appropriate for analysis.[42] We do not find

---

[41]Reuters reported that the boycott tarnished "the image of a country held up as a model of democracy in Africa," as turnout in the parliamentary election plunged "below 35%—less than half the more than 70% recorded in the February presidential election" (Reuters News 2007b). Commentators already noted in the immediate aftermath of the February election that the presidential campaign "has exposed cracks in the country's democratic fabric" (Reuters News 2007a). The only major observer mission appears to have been deployed by the Economic Community of West African States (ECOWAS), which sent 60 representatives to 9 out of 11 administrative regions (ECOWAS 2007a). ECOWAS issued a press release 2 days after the election that noted "isolated but serious incidents" but found the election "sufficiently free and transparent" (ECOWAS 2007b). We were unable to obtain the observation mission's report from ECOWAS headquarters in Abuja, Nigeria.

[42]In practice, this means that we include counts for Wade's main competition (Diouf in 2000, and Idrissa Seck and Ousmane Tanor Dieng, who received 15% and 14% of the official vote count, respectively, in 2007) as well as the number of registered voters for the election in 2000. Registration information was not available for the 2007 election. We do not include figures for total votes because they appear to have been computed automatically from other fields in the relevant spreadsheets. We analyze data for the decisive round of voting, that is, the run-off round in 2000 and the first round in 2007. Figures 9 and 10 exclude single- and double-digit counts because they are relatively prevalent in these data compared to the other election results we analyze. The patterns illustrated here persist if we include small counts, but significant deviations from the uniform distribution become more pronounced for the 2007 election.
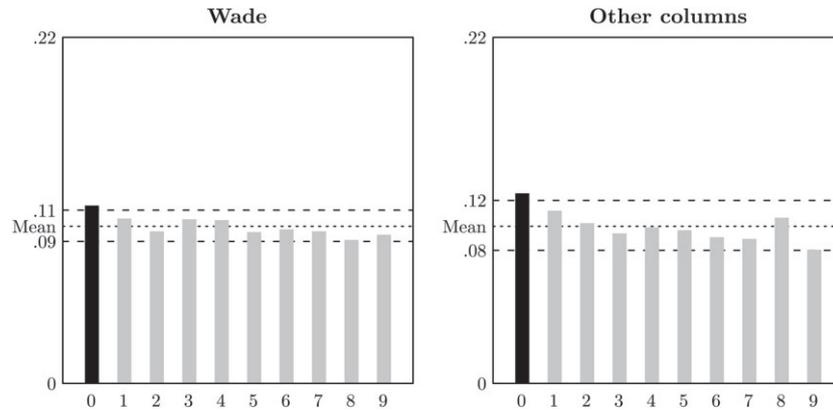
Fig. 10 Frequencies of last digits, Senegal 2007.

significant departures from equifrequent last digits for the reportedly clean election in 2000, whereas the results for 2007 show statistically significant deviations, just as we would expect.[43]

Particularly important for the purposes of validating our analysis of the data from Nigeria is the fact that we observe an overabundance of zeros in Senegal's allegedly fraudulent 2007 election. Zero is the most common trailing digit in this election, just as it was in the case of Nigeria. One could object that this fact reflects rounding or other nonfraudulent actions taken by election officials in West Africa. But this objection cannot easily stand in the face of the absence of excessive zeros in the reportedly clean 2000 election.[44] If overabundant zeros were really just the result of upstanding West African election officials going about their work in a fair and evenhanded but imprecise and expedient manner, then we should also find an excess of zeros in Senegal's uncontroversial 2000 election. This is not the case.[45]

5   Conclusion

This paper derived and applied a method to detect manipulation of electoral return sheets. We showed that we can expect the last digits of electoral results to occur with equal frequency given a wide range of distributional assumptions, and we then emphasized the fact that humans tend to be biased in the production of random numbers: They tend to select small digits, avoid repetition, and favor adjacent numerals. If we find that digit patterns deviate from our theoretical expectation in a way that reflects these biases, we suspect that a return sheet has been manipulated. We used data from Sweden, Nigeria, and Senegal to show that our approach is sensitive to known fraud but produces a null result in a nonfraudulent environment.[46]

The diagnostic tool developed in this paper is limited in that it detects only a specific type of fraud: the fraudulent (and nonalgorithmic) production of numbers in their entirety on vote returns. This means that election monitors should want to deploy this method in combination with other techniques. But allegations of fabricated vote returns are not rare; they are a common ingredient of the reports of election monitors,

---

[43]A chi-square test yields *p*-values of .45 and .80 for 2000 and .001 and .002 for 2007.

[44]Zero is the most frequent digit across candidate counts for the 2007 data, as was the case in Nigeria. This is not the case for the 2000 election, where zeros are not more frequent than any other numeral for the number of registered voters and votes for Wade.

[45]We have no reason to believe that an abundance of zeros in potentially fraudulent elections is a distinctly West African phenomenon, either. In the case of Chicago discussed in the supplementary materials that are available online, zero is the most common last numeral in the vote counts that we identify as suspicious (i.e., the returns for candidates backed by Chicago's then-Republican machine, Coolidge in 1924 and Hoover in 1928) but not otherwise (i.e., in the returns for their Democratic opponents, Davis in 1924 and Smith in 1928).

[46]We analyze an intermediate case, where fraud appears probable but has not been documented, in the supplementary materials available on the *Political Analysis* Web site, using data from Chicago for the 1924 and 1928 presidential elections that were generously shared with us by Kevin Corder. On the case of Chicago, see also Corder and Wolbrecht (2006).

and our method can be a powerful tool in hindering election fraud in general by making it more difficult for officials to rig elections by fabricating return sheets.[47]

When a new forensic tool like the one we describe here is adopted by election monitors, cheating officials may try to adapt and avoid detection by shifting to a different strategy of fraud. But this is almost certainly not costless for the official—why else engage in the wholesale fabrication of numbers in the first place?—and will therefore decrease the net benefit of electoral fraud.[48] Even if officials adapt and cheat in ways that cannot be detected by the tests described here, the tests can thus continue to serve the purpose of raising the cost of fraud.[49]

References

Agence France Presse. 2000. Mbeki congratulates Wade for winning Senegal election. Published March 22, 2000.

———. 2007. Catholic church joins critics of Nigeria poll amid fresh calls for re-run. Published April 25, 2007.

Al-Marzouki, Sanaa, Stephen Evans, Tom Marshall, and Ian Roberts. 2005. Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *British Medical Journal* 331:267–70.

Alvarez, R. Michael, Thad E. Hall, and Susan D. Hyde, eds. 2008. *Election fraud: Detecting and deterring electoral manipulation.* Washington, DC: The Brookings Institution.

Beber, Bernd, and Alexandra Scacco. 2011. Replication data for: What the numbers say: A digit-based test for election fraud. IQSS Dataverse Network [Distributor] V1 [Version]. http://hdl.handle.net/1902.1/17151 (accessed February 25, 2012).

Boland, Philip J., and Kevin Hutchinson. 2000. Student selection of random digits. *Statistician* 49:519–29.

Budescu, David V. 1987. A Markov model for generation of random binary sequences. *Journal of Experimental Psychology* 13:25–39.

Cantú, Francisco, and Sebastián M. Saiegh. 2011. Fraudulent democracy? An analysis of Argentina's *Infamous Decade* using supervised machine learning. *Political Analysis* 19:409–33.

Chapanis, Alphonse. 1953. Random-number guessing behavior. *American Psychologist* 8:332.

Corder, J. Kevin, and Christina Wolbrecht. 2006. Political context and the turnout of new women voters after suffrage. *Journal of Politics* 68:34–49.

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

Diekmann, Andreas. 2007. Not the first digit! Using Benford's law to detect fraudulent scientific data. *Journal of Applied Statistics* 34:321–9.

Dlugosz, Stephan, and Ulrich Müller-Funk. 2009. The value of the last digit: Statistical fraud detection with digit analysis. *Advances in Data Analysis and Classification* 3:281–90.

ECOWAS. 2007a. ECOWAS electoral observers in Senegal. Press release 11/2007, February 25, 2007.

———. 2007b. ECOWAS observers endorse presidential election in Senegal. Press release 13/2007, February 27, 2007.

European Union Election Observation Mission. 2003. Nigeria 2003: Final report on the national assembly, presidential, gubernatorial, and state houses of assembly elections. http://ec.europa.eu/external_relations/human_rights/eu_election_ass_observ/nigeria/rep03.pdf (accessed March 20, 2008).

Ichino, Nahomi, and Matthias Schündeln. 2011. Deterring or displacing electoral irregularities? Spillover effects of observers in a randomized field experiment in Ghana. Working paper.

Kevles, Daniel J. 1998. *The Baltimore case: A trial of politics, science, and character.* New York: W. W. Norton.

Kew, Darren. 1999. Democrazy: Dem go craze, O: Monitoring the 1999 elections. *Issue: A Journal of Opinion* 27:29–33.

———. 2004. The 2003 elections: Hardly credible, but acceptable. In *Crafting the new Nigeria: Confronting the challenges*, ed. Robert I. Rotberg, 139–73. Boulder, CO: Lynne Rienner.

Kuhl, J., and W. Schönpflug. 1974. Ordnungsverhalten, Lärmbelastung, und Persönlichkeit. *Psychological Research* 37:143–76.

Levin, Ines, Gabe Cohn, R. Michael Alvarez, and Peter C. Ordeshook. 2009. Detecting voter fraud in an electronic voting context: An analysis of the unlimited reelection vote in Venezuela. Online Proceedings of the Electronic Voting Technology Workshop.

Mebane, Walter R., Jr. 2006. Election forensics: Vote counts and Benford's law. Working paper.

---

[47]In principle, the tests we propose can be applied to election results at any level of aggregation, but there are two reasons why monitors may want to apply the tests to relatively disaggregated figures and pressure officials to release the necessary data: First, the fabrication of aggregate results can be constrained by figures released for lower levels of aggregation, which is not the case for fake counts at the initial level of tabulation. It is possible that national-level officials can make up aggregate counts and subsequently publish adjusted lower-level returns as needed, but the burden of having to reconcile numbers across levels of aggregation does not constrain officials at the bottom of the tabulation scheme in the first place. Second, the tests proposed here will be more powerful in a statistical sense when applied to disaggregated vote counts simply because the total number of observations will be larger.

[48]Watrin, Struffert, and Ullmann (2008) provide evidence from an experiment showing that the net benefit from tax evasion declines when a digit test based on Benford's law is applied to subjects.

[49]Alternatively, the test could be deployed clandestinely to prevent adaptation by election officials, although this may not be feasible if organizations engage repeatedly in monitoring and reporting.

———. 2008. Election forensics: The second-digit Benford law's test and recent American presidential elections. In *Election fraud: Detecting and deterring electoral manipulation,* eds. R. Michal Alvarez, Thad E. Hall, Susan D. Hyde, 162–81. Washington, DC: The Brookings Institution.

———. 2011. Comment on ''Benford's law and the detection of election fraud.'' *Political Analysis* 19:269–72.

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* 48:392–411.

Mosimann, James E., and Makarand V. Ratnaparkhi. 1996. Uniform occurrence of digits for folded and mixture distributions on finite intervals. *Communications in Statistics—Simulation and Computation* 25:481–506.

Mosimann, James E., Claire V. Wiseman, and Ruth E. Edelman. 1995. Data fabrication: Can people generate random digits? *Accountability in Research* 4:31–55.

Mosimann, James E., John E. Dahlberg, Nancy M. Davidian, and John W. Krueger. 2002. Terminal digits and the examination of questioned data. *Accountability in Research* 9:75–92.

Myagkov, Mikhail, Peter C. Ordeshook, and Dimitry Shakin. 2005. Fraud or fairytales: Russia and Ukraine's electoral experience. *Post-Soviet Affairs* 21:91–131.

Myagkov, Misha, Peter C. Ordeshook, and Dimitri Shakin. 2007. The disappearance of fraud: The forensics of Ukraine's 2006 parliamentary elections. *Post-Soviet Affairs* 23:218–39.

Myagkov, Mikhail, Peter C. Ordeshook, and Dimitri Shakin, eds. 2009. *The forensics of election fraud: Russia and Ukraine.* New York: Cambridge University Press.

Nickerson, Raymond S. 2002. Perception and production of randomness. *Psychological Review* 109:330–57.

Nigrini, Mark J. 1999. I've got your number: How a mathematical phenomenon can help CPAs uncover fraud and other irregularities. *Journal of Accountancy* 187:79–83.

Preece, D. A. 1981. Distributions of final digits in data. *Statistician* 30:31–60.

Rapoport, Amnon, and David V. Budescu. 1997. Randomization in individual choice behavior. *Psychological Review* 104:603–17.

Rath, Gustave J. 1966. Randomization by humans. *American Journal of Psychology* 79:97–103.

Reuters News. 2007a. Senegal court confirms Wade reelected president. Published March 11, 2007.

———. 2007b. Wade party wins Senegal poll: Foes threaten protest. Published June 7, 2007.

Schäfer, Christin, Jörg-Peter Schräpler, Klaus-Robert Müller, and Gert Wagner. 2004. Automatic identification of faked and fraudulent interviews by two different methods. German Institute for Economic Research Discussion Paper.

Smith, Jane, and Fiona Godlee. 2005. Investigating allegations of scientific misconduct: Journals can only do so much; institutions need to be willing to investigate. *British Medical Journal* 331:245–6.

The Economist. 2007. Nigeria: How to steal an election. Published April 18, 2007. http://www.economist.com/node/9032254 (accessed July 31, 2011).

Tversky, Amos, and Daniel Kahneman. 1972. Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3:430–54.

United Nations Development Programme. 2007. Human Development Report 2007/2008. http://hdr.undp.org/en/media/HDR_20072008_EN_Complete.pdf (accessed July 31, 2011).

Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane. Jr., Michael C. Herron, and Henry E. Brady. 2001. The butterfly did it: The aberrant vote for Buchanan in Palm Beach County, Florida. *American Political Science Review* 95:793–810.

Watrin, Christoph, Ralf Struffert, and Robert Ullmann. 2008. Benford's law: An instrument for selecting tax audit targets? *Review of Managerial Science* 2:219–37.

White, Caroline. 2005. Suspected research fraud: Difficulties of getting at the truth. *British Medical Journal* 331:245–6.

World Bank. 2011. World development indicators. http://databank.worldbank.org (accessed July 31, 2011).

Yule, G. Udny. 1927. On reading a scale. *Journal of the Royal Statistical Society* 90:570–87.