MEASURING CORRUPTION:

UNMASKING STRATEGIC DATA MANIPULATION

AND THEFT OF DEVELOPMENT AID

Jean Ensminger (California Institute of Technology)
Jetson Leder-Luis (Massachusetts Institute of Technology)

Draft. Please do not cite without permission.

Comments most welcome, please direct to:
Jean Ensminger: jensming@hss.caltech.edu

MEASURING CORRUPTION:

UNMASKING STRATEGIC DATA MANIPULATION*

JEAN ENSMINGER AND JETSON LEDER-LUIS

Abstract

Corruption is a pervasive and destructive problem that undermines economic development, democracy, and the effectiveness of aid. Efficient identification and measurement are key to finding policy solutions. We present new tests to detect fraudulent patterns in data and apply these methods to a World Bank project in Africa. Digit analysis exploits the fact that humanly produced data follow different patterns than naturally occurring data. Motivated by economic and psychological theories that guide predictions about fraudulent behavior, we develop new tests to expose strategic and profitable data fabrication built upon Benford's Law of natural digit distributions. Other new tests significantly expand the use of digit analysis to smaller sample sizes and uncover behavior consistent with attempts to subvert detection. An unusual aspect of this case is that we have external validity for our empirical findings. A forensic audit of the same project, conducted by the World Bank, revealed that 62 percent of project transactions were suspected fraudulent or questionable, and the distribution of these fraudulent patterns across districts corresponds with the results of our method. Our methods improve upon existing anti-corruption techniques in two ways: their application requires little cooperation from potentially corrupt insiders, and they provide a scalable, real time, and cost-effective method to monitor suspected fraud in public and private expenditures. *JEL* Codes: D73, O22, H83, C49

* We particularly thank Avinash Dixit, Esther Duflo, Ben Gillen, Jonathan Katz, Antonio Rangel, and Robert Sherman for advice and support at critical junctures of this project. In addition, we have benefitted from the comments of colleagues and seminar participants at presentations of earlier drafts of this paper at the ASSA meetings, Massachusetts Institute of Technology, California Institute of Technology, New York University, Center for Global Development, Duke University, and Oxford University. We thank the California Institute of Technology for generous funding of this research.

MEASURING CORRUPTION:

UNMASKING STRATEGIC DATA MANIPULATION

JEAN ENSMINGER AND JETSON LEDER-LUIS

I. INTRODUCTION

Corruption is an insidious and widespread problem, undermining economic development, democracy, and even national security. Banerjee, Hanna, and Mullainathan (2012) note that the study of corruption is fundamentally impeded by measurement challenges, because those committing corruption attempt to hide it. In order to combat corruption, we must be able to find and measure it using a method that is scalable and functional in highly corrupt environments. In this paper we develop new digit analysis methods that exploit irregular patterns in the digits of humanly generated numbers. Our tests are derived from two principles: that fraudulent data manipulation occurs as part of an economic decision to commit crime, and that humans are poor generators of random numbers, as evidenced by psychology research. We validate our new tests using data from an African World Bank development project for which we have an existing forensic audit with consistent results. Our method allows for cost-effective scaling of monitoring efforts while requiring minimal cooperation from those who may be implicated in corruption.

Our work relates to theoretical literature on both the economics of committing and monitoring corruption. We motivate our statistical tests by considering the decision process of a bureaucrat choosing to commit fraud. This decision depends on the incentives of the bureaucrat to maximize rents as well as the costs of being caught (Becker 1968, Besley and McLaren 1993, Schleifer and Vishny 1993). On the monitoring side, a principal-agent approach describes how pervasive corruption increases monitoring costs because there are fewer whistleblowing agents (Lui 1986). As a consequence, those monitoring corruption must be cautious with use of whistleblower information, and should increase monitoring of those not directly suspected of corruption (Chassang and Padró i Miquel 2014). Our development of a low cost monitoring technique addresses these problems by providing methods for more widespread, continuous monitoring.

A variety of techniques have been successfully employed in the fight against corruption. Audits have proven to be a useful tool in identifying, measuring, and deterring corruption. Ferraz and Finan (2008) show that randomized audits affect the electability of corrupt officials in Brazil. Avis, Ferraz and Finan (2016) show that these same audits deter subsequent corrupt behavior. Olken (2007) demonstrates that a rise in Indonesian government audits from 4 percent to 100 percent decreases missing expenditures in road projects by 8.5 percent. Despite their effectiveness, audits are difficult to implement, as they are labor intensive, use highly trained personnel, and require cooperation from individuals who may be implicated in the corruption, including government auditors themselves (Duflo, Greenstone, Pande and Ryan 2013). Similarly, monitoring efforts that track expenditures (e.g. Reinikka and Svensson 2006)) suffer from issues of scale and the need for cooperation from those who may be complicit in fraud.

A number of creative studies have pioneered methods for identifying and measuring corruption that require no cooperation from the subjects being monitored. Satellite data have been used to track illegal logging (Burgess, Hansen, Olken, Potapov and Sieber 2012); stock market fluctuations have been used to measure financial returns to political influence in Indonesia (Fisman, Fisman, Galef, Khurana and Wang 2006, Fisman 2001); and people have directly observed bribery (Mcmillan and Zoido 2004, Olken and Barron 2009). However, all of these methods have limited application, either because they are specific to one issue or because they would be cost prohibitive to scale.

Digit analysis relies only on reported data, and thus is highly scalable, cost-effective, and can be deployed contemporaneously as government, private, or donor projects roll out. Digit analysis explicitly detects patterns in data that reflect an attempt to manipulate the data. It can be extremely effective in targeting audits toward the most problematic areas, thus achieving cost efficiencies. And, as it does not require cooperation from potentially corrupt officials, it can be effective in systemically corrupt countries such as Kenya, where powerful cartels control large parts of the government (Lindijer 2016).

The remainder of this paper is organized as follows. Section II motivates our digit analysis tests with a discussion of the economics of data manipulation and an overview of the mathematical principles that govern digit distributions. Section III describes our Dataset, and Section IV presents our statistical tests and their results. Section V compares our results to the World Bank forensic audit of the same project, and section VI concludes.

II. MOTIVATION AND THEORY

We begin by examining the problem of a bureaucrat's decision to accurately report or to fabricate data when tasked with producing expenditure reports, such as those common to development aid projects.

Using a set of receipts, an honest bureaucrat has the opportunity to produce a legitimate report. In this case, the bureaucrat will sum receipts to produce line item totals, and will also accurately report information such as participants or beneficiaries in these aid projects. These data follow the digit patterns of natural data, described below, as they accurately reflect the data without human interference. Across different geographic regions of this World Bank project, we would expect similar patterns in the financial data when reporting is conducted honestly in each region. Data that deviate from our expected distributions should not show similar patterns across the expenditure and participant data. If a bureaucrat struggles with sloppy bookkeeping, we might expect an overuse of rounding, but not in a way that systematically inflates values or shows human patterns consistent with strategic data manipulation.

There are many reasons why a bureaucrat may engage in dishonest reporting. Bureaucrats have an incentive to falsify expenditure data and embezzle, both for personal gain as well as to satisfy kickback demands from higher ups. Detection of dishonest reporting is costly; it can have career consequences for the embezzler, but also may require a payoff to the auditor or higher-up who detects this fraud. Furthermore, data manipulation is costly in both time and effort. Therefore, dishonest bureaucrats cheat in a way that reflects rational criminal behavior, as per Becker and Stigler (1974). We can expect the following attempts to manipulate data: 1) padding high digits in more valuable digit places, 2) more data tampering in response to greater incentives to steal, 3) attempts to produce data that appear random to subvert detection, and 4) lower effort expended upon data generation that is less likely to be monitored

The data generation process described above yields predictions that motivate statistical testing. While data produced legitimately should be composed of digits which follow the distribution of un-tampered data, fraudulent data will show patterns consistent with human generation of digits, as well as profitable deviations, an intent to hide their fraud, and patterns of human tampering that are similar between expenditure data and participant data.

Digit analysis is the comparison of the expected and observed frequency of the digits 0 through 9 within a data set. Humanly produced data follow different patterns than naturally occurring data. Digit analysis has been used in auditing (Amiram, Bozanic and Rouen 2015, Durtschi, Hillison and Pacini 2004, Nigrini 2012, Nigrini and Mittermaier 1997), election fraud

(Alvarez, Hall and Hyde 2008, Beber and Scacco 2012), campaign finance fraud (Cho and Gaines 2012), scientific data fabrication (Diekmann 2007), IMF data monitoring (Michalski and Stoltz 2013) and the monitoring of enumerator integrity during survey research (Bredl, Winker and Kötschau 2012, Schräpler 2011).

A very simple test for data manipulation is the analysis of terminal digits. In naturally occurring data, terminal digits should be uniformly distributed; there is no reason more date should end with a 4 instead of a 3. When humans produce data, they fail to follow the uniform distribution for the last digit, instead encoding preferences for certain digits (Boland and Hutchinson, Nigrini). We implement one test based on uniformly distributed terminal digits below.

Many of our tests are founded upon Benford's Law, which gives the expected distribution of digits in other digit places of financial data. Benford's Law describes the distribution of digits in many naturally occurring circumstances, including financial data. Benford's Law is given mathematically by (Hill 1995):

$$P(D_1 = d_1, \dots, D_k = d_k) = \log_{10}\left(1 + \frac{1}{\sum_{i=1}^{k} d_i \times 10^{k-i}}\right)$$

We have, for example:

$$P(D_1 = 4, D_2 = 5, D_3 = 2) = \log_{10}\left(1 + \frac{1}{452}\right)$$

In the first digit place, this produces an expected frequency of 30.1 percent of digit 1 and 4.6 percent of digit 9. In later digit places, this curve flattens, and by the 4th digit place the distribution is nearly identical to the uniform distribution, with expected frequency 10.01 percent of digit 1 and 9.98 percent frequency of digit 9 (Hill 1995, Nigrini and Mittermaier 1997). The full digit-by-digit place table of expected frequencies under Benford's Law is given in Table 1. Datasets known to follow Benford's Law include financial data and population data, but also everything from scientific coefficients to baseball statistics (Amiram, Bozanic and Rouen 2015, Diekmann 2007, Hill 1995, Nigrini and Mittermaier 1997).

The intuition behind Benford's Law is revealed if one imagines it as a piling effect: increasing a first digit 1 to a digit 2 requires a 100 percent increase, while increase from a first digit of 8 to a first digit of 9 requires a 12 percent increase (Nigrini and Mittermaier 1997). Furthermore, Benford's Law arises from data drawn as random samples from random distributions (Hill 1995). Because numbers repeatedly multiplied or divided will limit to the Benford distribution, financial data can be expected to follow this natural phenomenon (Boyle 1994).

The appropriateness of Benford's Law for use in our data set is confirmed by the conformance of the first digits to the Benford distribution, as we show later. The nature of our expenditure data, which are based upon sums of numerous receipts that in turn include sums and multiplication of price times quantity, provides a theoretical basis for why we can expect Benford's Law to be the distribution of interest. In our analysis, we consistently perform robustness checks by comparing our observed distributions to both the Benford and the uniform distributions; the statistical significance under the uniform distribution is even greater than those reported here. Finally, regardless of Benford's Law, tests of later digit places, particularly last digits, should be uniformly distributed under most conditions.[1]

[1] In the study of elections, the use of Benford's Law has been contested based on concerns over the distributions of data that produce voting counts Beber, Bernd, and Alexandra Scacco, "What the numbers say: A digit-based test for election fraud," *Political Analysis*, 20 (2012), 211-234, Deckert, Joseph, Mikhail Myagkov, and Peter Ordeshook, "Benford's Law and the detection of election fraud," ibid.19 (2011), 245-268, Walter R. Mebane, Jr., "Comment on "Benford's Law and the detection of election fraud"," ibid.269-272.. However, these criticisms do not extend to our financial dataset or individual participant counts, both of which come from distributions that can be expected to conform to Benford's Law. Specific auditing guidelines over which types of data conform to Benford's Law

In contrast, studies of individuals producing numbers consistently show divergent patterns of human digit preferences. In experiments where students are asked to make up strings of 25 digits, their results follow neither the Benford distribution nor the uniform distribution (Boland and Hutchinson 2000). The patterns produced by the subjects varied greatly, with individuals exhibiting different preferences for certain digits. Other experiments have shown similar results of individual digit preferences, confirming the inability of humans to produce random digits (Chapanis 1995, Rath 1966).

A deeper contextual understanding of the specific digit preferences of Africans comes from an overview of African census data, where statisticians discuss a phenomenon known as age heaping, wherein self-reported demographic records show a preference for certain ages. Many Africans of older generations do not know their exact age, and their responses to census takers represent their best approximation. This is an example of humanly generated data that shows extreme digit preferences. Unlike the subjects in the experiments cited earlier, the individuals producing these data have similar numeracy to the subjects in this analysis. Among the African censuses, we see a strong preference for the digits 0 and 5, with secondary strong preferences for 2 and 8, and disuse of 1 and 9 (Nagi, Stockwell and Snavley 1973, UN Economic and Social Council Economic Comission for Africa 1986). One potential explanation is that individuals are most comfortable rounding to 0s and 5s, and when they know their age more precisely, they choose 2 and 8 as the mid-point, revealing a preference for even numbers (over 3 and 7) and avoiding 4 and 6 and 1 and 9 because they flank 5 and 0, which are the more common defaults.

Regardless of the exact distributions from which individuals committing fraud draw their made up numbers, experiments indicate that there is virtually no chance that they will conform to

---

includes these types of data Durtschi, Cindy, William Hillison, and Carl Pacini, "The effective use of Benford's Law to assist in detecting fraud in accounting data," *Journal of Forensic Accounting*, V (2004), 17-34..

either the Benford or the uniform distributions. This motivates the method of digit analysis, which compares observed frequencies of digits in financial or other datasets to the appropriate theoretical distribution.

Through analysis of the digit distributions in the participant and expenditure data, and by comparison across geographic regions and different types of expenditure, we can observe signals of human tampering that indicate fraudulent reporting. Uniquely, this measurement method detects efforts to report fraudulently without requiring cooperation from a corrupt bureaucrat. We now turn to a discussion of the data we use to exhibit our methodological contribution.

III. DATASET

We analyze data from the Kenyan Arid Lands Resource Management Project (World Bank 2003). This World Bank project ran from 1993 to 2010, eventually serving 28 arid and semi-arid districts over 75 percent of Kenya's land area. The project spent US$224 million on development targeting the most impoverished people in the heavily drought-prone regions of Kenya. It funded small infrastructure (schools, dispensaries, and water systems), income generating activities, drought and natural resource initiatives, and training exercises for villagers.

The data used in these analyses are from the original 11 arid districts that received funds from the project. Most of these 11 districts practice a predominantly pastoral lifestyle: that is, they are dependent to a great degree upon livestock economies of cattle, sheep, goats, and camels. The districts are among the poorest in Kenya, remote from centers of power, poorly educated, and sparsely supplied with infrastructure (roads, schools, health services, access to clean water, and electricity). The ecologies are all arid and prone to frequent droughts; agriculture is viable only in relatively small parts of each district.

The expenditure and participant data used in these analyses were culled from electronic project reports produced in each of the 11 districts. Interview data from those districts indicate that usually only 1 or 2 individuals were involved in entering the data for each sector of the project. Reports were typically organized around the various project components (natural resources and drought management, community driven development, and support for local development) and sectors (civil works, goods and equipment, training). Officers in the districts had considerable latitude over the magnitude of expenditures within budget categories, with the exception of community driven development projects. These expenditures were grants to communities subject to caps, and as such were not appropriate for Benford's Law data analysis, and are removed. In addition to the expenditure data, we use data on the number of participants who attended training exercises and benefitted from projects.

In 2009, the World Bank's Integrity Vice Presidency (INT) began a forensic audit of the project that lasted 2 years and culminated in a public report (World Bank 2011). Auditors sampled 2 years' worth of receipts for 7 districts, 5 of which were arid districts examined in this analysis. They examined 28,000 transactions and classified 62 percent of them as suspected fraudulent or questionable. The auditors worked from actual project receipts and supporting documents such as cashbooks and bank statements and travelled to the districts to verify the legitimacy of suspicious transactions. We conduct digit analysis on the reported total expenditures for each of these transactions, as they appear in the district annual reports. Our analysis is on the reported expenditures per activity, while the forensic auditors investigated the underlying individual receipts for the same transactions.

We make use of 2 characteristics in the structure of our dataset. First, we have data from 11 arid districts with similar demographics, livelihoods, and ecological conditions, reporting on similar activities, and operating under the same project rules. We proceed from the null hypotheses that digit distributions will be similar across districts. Second, we have data on the number of participants in hundreds of training exercises and activities. Deviations from theoretical distributions that appear in both the expenditure and the participant datasets are strongly indicative of human tampering. They allow us to rule out a legitimate digit aberration that might be unique to these particular expenditure patterns, because the same irregular digit pattern should not also show up in a count of people who responded to an open invitation for a training exercise.

IV. DIGIT TESTS AND RESULTS

Our discussion now turns to the digit tests. We begin with a number of common tests from the literature, noting modifications that are appropriate for datasets such as ours. Next, we present 4 completely new tests that expand the power of digit analysis, particularly by aiding the analysis of small sample sizes, and by finding evidence of strategic data manipulation that speaks to intent to defraud. We then compare the overall findings from the district level results of our tests, tabulating how many of the 12 tests each district failed, and comparing these results to the level of suspected fraud for the same districts from the World Bank's forensic audit.

IV.A. First Digit

Our first test is for conformance to the Benford distribution in the first digit place of the expenditure data, where we expect digits to follow (Hill 1995):

$$\text{P(First Digit} = d) = \log_{10}\left(1 + \frac{1}{d}\right)$$

Benford's Law predicts an unequal distribution of digits in the first digit place, with 30 percent 1s and 4.6 percent 9s. Figure 1A plots this distribution as a solid line and shows the conformance of the first digits of the expenditure data to Benford's Law. Data from the full sample are not statistically significantly different from the expected distribution ($p = 0.089$) under a chi-square test. This supports the hypothesis that Benford's Law is the appropriate theoretical distribution for our dataset. Importantly, this does not indicate that the data are legitimate, as pooled fraudulent data may cancel out different individual signatures and replicate Benford's Law (Diekmann 2007). This becomes evident when we look at the data from individual districts where the reports were constructed. Figure 1B shows the first digits from Ijara district, with $p = 2.3 \times 10^{-13}$. Ijara District uses the digit 2 in the first digit place almost twice as often as predicted. Seven of our 11 districts are significantly different from Benford's Law at the $p < 0.005$ level. This significance level is chosen in line with a Bonferroni correction, described later.

[Figure 1AB here]

IV.B. High Second Digits

Tests of second digits are a common practice in digit analysis, particularly because first digit manipulation is not always available to those perpetrating fraud (da Silva and Carreira 2013, Nigrini 2012, Nigrini and Mittermaier 1997). For this reason, we hypothesize that individuals will pad the second digit. Thus we explicitly test for an overuse of the digits 8 and 9 in the second digit place. Omitting 0s and 5s as discussed in the preceding section, Benford's Law predicts 11.2 percent 8s and 10.8 percent 9s. Combined, this gives a 22 percent probability of finding these two highest digits. For each district, we analyze the second digit for overuse of these digits, using a binomial test to compare the observed number of 8s and 9s to the Benford expected value of 22 percent. Figure 2 shows the results. Two districts, Baringo and Moyale, significantly overuse these digits in the second digit place at the $p < 0.005$ level. Furthermore, the sample of all districts together is statistically significant.

[Figure 2 here]

IV.C. Last Digits

Individuals tampering with data may not have complete control over the leading digit, or may avoid changing it to subvert detection. Consistent with the literature on both forensic auditing and election fraud, we analyze the terminal digits for conformance to the uniform distribution, where each digit is equally likely (Beber and Scacco 2012, Nigrini and Mittermaier 1997). Benford's Law limits to the uniform distribution for digit places beyond the fourth.

Importantly, this dataset showed an excessive level of rounding, discussed later. So as not to doubly penalize rounding, we eliminate 0s and 5s from the last digit. With 8 digits (1,2,3,4,6,7,8,9), the uniform distribution gives an expectation of 12.5 percent (1/8) per digit. We compare our observed frequencies to this modified uniform distribution with a chi square test.

Figures 3A and 3B show the distribution of the final digits of the entire dataset for the expenditure and participant results respectively. The expenditure data are highly significant when compared to the uniform distribution, $p = 1.5 \times 10^{-9}$ with $n = 851$. Among participant data, we use the terminal digits from the aggregation of the male and female reported beneficiaries lumped together. We see even stronger significance, $p = 7.04 \times 10^{-26}$, with $n = 5850$.

[Figure 3AB here]

Overall, the final digits test is the most widely used and indisputable test of human tampering. In these data we consistently see strong indicators of fraud due to the divergence from the uniform distribution. These data exhibit a strong preference for the digits 2 and 8, and an underrepresentation of digits 1 and 9, in both the expenditure and the participant data. This pattern is exactly what appears in the African census data (Nagi, Stockwell and Snavley 1973), where we know that data were humanly produced. Due to heavy rounding, our sample size for the last digit analysis of expenditure data (n = 851) was insufficient to allow partitioning by district, but for the participant data (n = 5850), 8 out of 11 districts were significantly different from Benford's Law at the $p < 0.005$ level.

IV.D. Rounding, Repeats, and Sector Effects

Three tests further uncover patterns indicative of human tampering, as evidenced by substantial variation across districts when there is no plausible naturally occurring explanation for such variation. It is common for auditors to look for both high levels of rounded and repeated data, and these are often viewed as potential evidence of human tampering (Nigrini 2012, Nigrini and Mittermaier 1997). In the absence of theoretically acceptable levels of rounding and repeating, we compare districts to each other, as there is no reason to expect differences among them.

The Kenyan shilling was 66 to US$1 in 2008. Its value was low enough that many receipt data would legitimately show high levels of 0s and 5s in the terminal digit places. However, these expenditure data represent sums of many receipts; it takes only one receipt ending in a non-0 or 5 to create a different terminal digit for the transaction.

To address rounding, we count rounded digits rather than rounded line items, tallying the number of trailing 0s (0, 00, 000, etc.) or digits in terminal strings of 5, 50, or 500, as a fraction of the total digits in the district dataset. For example: the number 30,000 has 4 rounded digits; the number 12,350 has 2 rounded digits; and the number 11,371 has 0 rounded digits. Rather than indicating individual line items, counting rounded digits is a more sensitive indicator because it penalizes use of numbers such as 10,000 (4 rounded digits) more than the use of a number such as 10,600 (2 rounded digits). We compute the percentage of rounded digits for each district. Figure 4A shows the percentage of rounded digits by district, with the crosshatched districts in the top quartile of rounding. While we don't know the empirically correct level of rounding that one should observe in a dataset, there is good reason to expect that retailers across the Kenyan arid lands districts practiced the same rates of rounding. In the absence of an expected level of rounding, we flag those three districts, roughly the top quartile, that round most heavily, which is well over twice that of the lowest rounding district.

Figure 4B shows a similar result for the percentage of line items that repeat exactly. We define an exact repeat to mean an expenditure matching year, district, sector, and expenditure value. Note that duplicate entries for the same project were removed from the dataset; repeating numbers refer to the use of identical expenditure amounts for completely different projects. A specific example from the Tana District Report of 2003-6 illustrates this (Republic of Kenya 2006). On page 49 we find 8 training exercises listed that took place in different villages for three weeks, each from March 5-27. The district had neither enough vehicles, nor enough training staff to run 8 simultaneous trainings, which is a red flag in itself. The cost of these 8 trainings in Kenyan Shillings is listed as follows: 279,720; 231,420; 238,320; 249,447 (twice), and 245,392 (3 times). Trainings are the summed costs of the 4-5 trainers' and 1 driver's per diem allowances (at different rates) for the number of days they are away from station, the cost of fuel to the destination, stationary for the seminar, food for the trainers, and 100KSH per day, per trainee, for food costs. The number of trainees for each of these seminars is listed, and they range from 51 to 172. The expenses reported do not track the estimated food costs, as one would expect. Exactly repeated numbers are a red flag for auditors (Nigrini and Mittermaier 1997). Repeated values reflect low-effort data fabrication and may reflect the lack of scrutiny the project staff expected on training expenditures, which are hard to verify.

[Figure 4AB here]

As before, we indicate the top three districts that most heavily repeat numbers; for example, Baringo approaches 50 percent, while Turkana has about 5 percent. Although the empirically appropriate level of repeating is unknown, given that Turkana fails the second fewest digit tests overall, and Baringo is the second highest offender, one is inclined to assume that Baringo and the other high repeaters are the aberrant districts.

There is some overlap between these tests, as the number of repeated data may be inflated by the presence of a lot of rounded numbers. But this is not always what is driving repeats, as we see for Baringo district, which has the highest percentage of exactly repeated numbers, and has the lowest number of rounded digits. Therefore, these 2 tests do pick up independent evidence of human generation.

Auditing and economic theory indicate that individuals are more likely to cheat when there is a lower risk of detection (Becker 1968). Training, travel, fuel, and vehicle maintenance (transport) provide greater opportunities for individuals to conduct fraud when compared to civil works projects or the purchase of goods and equipment, because the latter provide physical evidence of the spending, while the former do not. For example, tracking down nomads who were reported as present for a training exercise in a village on the border of Kenya and Somalia two years prior to an audit is all but impossible. Similarly, fuel can be diverted to private vehicles while leaving no trace. Therefore, we predict that individuals fabricating data may do so with less attention to deception in these sectors. We look for evidence of this in a greater incidence of repeated numbers. We plot percentage of repeated line items, which match year, district, sector, and amount, for each of the districts by sector. Figure 5 shows this result. We crosshatch those districts that have three times the number of repeats in training and transport as compared to the average number of repeats in civil works and goods and equipment. Six of 11 districts and the all district test fail, but Turkana District, which has a low level of repeated data and less repeating in training than for the average of the other two sectors, provides evidence that there is no structural reason for there to be more repeated data in training and transport.

[Figure 5 here]

IV.E. Digit Pairs

Underuse of digit pairs, e.g. 11, 22…99, is a common feature of humanly produced data (Boland and Hutchinson 2000, Chapanis 1995). Other applications of digit analysis examine the last two digits (Nigrini 2012) or explicitly test for digit pairs (Beber and Scacco 2012). We propose a slightly modified test for use in examining terminal digit pairs in expenditure data due to confounding by rounding. Rather than counting digit pairs as a fraction of all data, we exclude 0s from the denominator (e.g. 10). We then compare the observed number of digit pairs against the expected proportion using a binomial test, where the number of trials is the total combination of terminal digits observed. Having controlled for rounding, and contrary to the existing literature, we find no evidence of underutilizing digit pairs in the expenditure data when we examine all districts together. However, this may be an aggregation problem, as Ijara District significantly overuses repeated pairs, with $p = 5.6 \times 10^{-5}$ in a binomial test. The removal of all zeros is nonstandard in conducting this test, and leaves us with insufficient sample size to run this test on most districts due to diminished n. However, we suggest that omitting zeros is a more appropriate test when the currency unit has low value, as is increasingly the case for the penny in the U.S.

We also run the binomial digit pair test on the participant data. These data most typically record the number of women and men (listed separately) who showed up in response to an open invitation to all villagers to appear for a training exercise in their own village; these were often remote pastoral-nomadic locations. There was no way for the trainers to know in advance how many people would attend. Participants in training exercises were supposed to sign-in each day. Among the participant data, we expect a uniform distribution of pairs, 9 of 99 pairs, as we omit the pair 00. Unlike expenditure data, there is no legitimate reason for excess single terminal 0s, and we do not perform the adjustment above. To avoid use of first digits, we use participant data only if it has 3 or more digit places. This test is performed on the sum of male and female participants, not on the individual gender breakout data. A digit pair analysis of participant data is shown in Figure 6. Six of the 11 districts significantly underuse final digits pairs in the participant data at $p < 0.005$ significance, as does the combined sample of all districts ($p = 1.4 \times 10^{-9}$).

[Figure 6 here]

We turn now to new tests that expand the use of Benford's Law to accommodate small samples and expose strategic data manipulation, including padding high numbers in high digit places.

IV.F. All Digit Places Beyond the First

Benford's Law tests often require disaggregation of data to expose individual signatures (digit preferences) that may cancel out when data from multiple agents are aggregated. This often puts pressure on sample sizes. We address this constraint with a new test that uses Benford's Law to simultaneously analyze all but the first digit place. The concurrent analysis of all digit places beyond the first is a substantially more powerful test than the individual digit place tests (first, second, and last) common to digit analysis (Beber and Scacco 2012, Diekmann 2007, Nigrini and Mittermaier 1997). When testing individual digit places, an analyst has one data point per line item number. By simultaneously analyzing multiple digit places, this test increases sample size for statistical testing and is beneficial in the analysis of small samples. We eliminate the first digit from this test because there is often less ability to tamper with the first digit place. We then use a two-way chi square test to compare the contingency table of all digit places beyond the first against the generalized Benford distribution using a chi-square test.

Figures 7A and 7B present the data of all digit places except the first for expenditure (7A) and participant data (7B), projected onto one axis for visualization. Among the expenditure data for all districts in Figure 7A, we see a strong preference for digits 2 and 8, underreporting of 1 and 9, and overall non-conformance to the expected Benford-weighted distribution ($p = 3.9 \times 10^{-15}$). Strikingly, these same digit patterns appear in the participant data (Figure 2B), and the result for all district data combined is again highly significant ($p = 5.7 \times 10^{-51}$). This pattern is consistent with the humanly generated African census pattern described earlier. In 8 our 11 districts we reject the null hypothesis that all digit places conform to Benford's Law for the expenditure data and for 8 out of 11 districts for the participant data at the $p < 0.005$ level.

[Figure 7AB here]

The lack of conformance to the expected distribution, consistency with known humanly generated data from African census studies, and similar patterns across both expenditure and participant data, are strong indicators that these data have been tampered with.

IV.G. Year Effects

We take advantage of the extra power afforded by use of multiple digit places simultaneously to partition our data by project year. This test is designed to detect potential fraud in 2007, when Kenya held national elections that increased incentives to embezzle money for political campaigns. We look for padding of high digit numbers by project year by analyzing the proportion of high digits (6, 7, 8, and 9) in all digit places beyond the first. We conduct a chi-square test on the contingency table of high and low digits, where the null hypothesis is that each of the two bins per digit place should follow Benford's Law, summing the probabilities of 6, 7, 8, 9 and 1, 2, 3, 4, respectively. As before, we project this contingency table onto one axis for visualization. As we see in Figure 8, while all other years slightly underused high digits on average, in 2007 (the election year) there was a statistically significant overuse of high digits ($p = 6.5 \times 10^{-6}$).

[Figure 8 here]

IV.H. Unpacking Rounded Numbers

Project staff had an incentive to inflate the number of participants in training activities because they claimed food expenses for each participant. The authors of the annual district reports also had reason to expect that participant data would not be as carefully scrutinized as expenditure data. First, the impact of participants on expenditures was obscured because it was only one component of the full costs of a single training exercise, and second, training exercises in remote villages are notoriously difficult to verify. With the threat of oversight reduced, we speculate that less effort was devoted to covering up data fabrication.

We further surmise that officers fabricating participant data may have begun with an embezzlement target in mind, which they converted to a round number of participants. This total number of participants was then split into males and females as was required for reporting. Therefore, we expected greater indicators of data fabrication when the total number of participants was a round number. To test this, we analyze the distribution of all but first digits of numbers of total participants (males and females) when their sum ends in a 0 versus a non-0 digit. We perform a chi-square test on the contingency table of digits in digit places beyond the first, versus Benford's Law. Theoretically, the breakout of participant data by gender should show statistically identical digit distributions between these conditions. However, we see a much higher instance of 2s and 8s and low incidence of 1s and 9s when the gender specific data come from a pooled number that ends in 0 (Figure 9A). This is indicative of human data fabrication. There is still evidence of human generation in the data when the gender total is not round, Figure 9B ($p = 1.9 \times 10^{-6}$), but the statistical significance is considerably higher in the rounded data, Figure 9A ($p = 2.6 \times 10^{-64}$), in the sample of all districts. We reject the null hypothesis that the data are Benford conforming for 8 out of 11 districts at the $p < 0.005$ level.

[Figure 9 here]

IV.I. Value of Digit Place with Monte Carlo Simulation

Our final new test reveals evidence of highly strategic fraudulent behavior. We identify padding of expenditures by measuring overuse of high digits based on the monetary value of the digit place. We hypothesize that individuals fabricating data do so strategically and therefore place additional high digits in the more valuable digit places.

We align the data by the value of the digit from the right (e.g. 1s, 10s, 100s place), rather than its position from the left (e.g. 1st digit, 2nd digit), as is common in Benford's Law analyses.

For each digit place from the right, we compute the expected mean under Benford's Law and subtract it from the observed mean. This is the difference of means statistics, for which a positive value indicates a mean greater than the expected mean under Benford's Law. We then perform a Monte Carlo simulation of 100,000 Benford-distributed datasets, and compare the difference-of-means statistic of the project data to the simulated data and find the probability of observing our results under the Benford distribution.

Figure 10 shows the project data by sector against the Benford expected distribution. The 0 line indicates the Benford mean; anything above the line represents an overuse of high digits, and anything below the line represents an underuse. The project data in the 10,000s place exceeded 100 percent of the 100,000 simulated Benford-conforming datasets ($p = 1.0 \times 10^{-5}$). We also see a significantly high mean ($p = 2.3 \times 10^{-4}$) in the thousands place. At the district level there is statistically significant evidence of padding in the 10,000's place for 8 of 11 districts; 10,000 Kenyan shillings was worth approximately US$150 in 2007. An interesting finding in Figure 10, which corroborates the strategic placement of digits, is the decline in the use of high digits as one goes from the 10,000s to the 1,000s, 100s, 10s, and 1s places among the pooled sector data, represented by the black bars. This is consistent with a strategy of padding extra high digits in the high value places and compensating by *underutilizing* high numbers in the low digit places. The human data generators may have been trying to avoid detection from an auditor or supervisor, who might otherwise have noticed the overuse of high numbers.

[Figure 10 here]

V. COMPARING DIGIT ANALYSIS TO THE WORLD BANK FORENSIC AUDIT

Table 2 compiles the results of 12 tests for each district. To correct for type 1 error due to the number of tests we ran, we perform a Bonferroni correction on the 9 tests for which we have *p*-values, choosing a significance level of 0.005 to account for multiple hypothesis testing. For the rounding and repeating tests where districts are compared to each other in the absence of a theoretical measure (Figures 4A and 4B), the top quartile of districts, 3 of 11, are flagged. For the sector effects (Figure 5), we mark those districts for which training and transport exhibit more than triple the data repeating of the other sectors. These 12 tests mostly avoid overlap and pinpoint different aspects of fraudulent data tampering. In the bottom row we sum the number of failed tests by district, which ranges from 3 to 9 out of 12.

[Table 2 here]

The existence of an extensive forensic audit for this project provides us with a measure of external validity for our digit analysis (World Bank 2011). In Table 3 we compare the results of our digit analyses by district to the results of the World Bank auditors. The World Bank audit found that 4 of the 5 districts for which we have both digit and audit results had 62-75 percent suspected fraudulent or questionable expenditures; in our digit analysis we rejected the null hypotheses for those same 4 districts in 6 to 8 of our 12 digit tests. The remaining district, Tana, had considerably lower levels of suspected fraud than other districts (44 percent), and we rejected the null on 3 of our 12 digit tests.

[Table 3 here]

We also found significant digit violations in all of the unaudited districts, which is consistent with the conclusions of the auditors that these problems were systemic throughout all sectors and all districts of the project. Of the remaining 6 districts that were not audited by the World Bank, we see that 2 (Baringo and Mandera) have the highest number of digit analysis violations (8 and 9) in our sample. This underscores the potential gains of using digit analysis as a diagnostic prior to an audit to target where to use costly auditing techniques.

VI. CONCLUSION

We present new methods to combat corruption. The struggle for solutions to corruption has been stymied in part by a lack of tools to find and measure the problem. In this paper we provide the first application of digit analysis to identify suspicious patterns in development aid data. We develop new digit tests and apply our methods to data from a large World Bank project, the Kenyan Arid Land's Resource Management Project. A forensic audit has been conducted on the same project, which allows us to externally validate our method and our results.

Digit analysis exploits the fact that humanly produced data follow different patterns than naturally occurring data. We modify many tests from existing digit analysis and develop 4 new tests that expand our capability to detect data tampering and strategic avoidance of detection. Our methods exploit the multi-dimensionality of our dataset. We examine data from 11 different project offices that should follow similar distributions because of the commonality of their circumstances and their project mandates. We uncover vastly different digit distributions consistent with individual signatures of humanly generated data. In addition, we find the same aberrant patterns across data from financial expenditures and sums of participants in training exercises. These digit patterns have been demonstrated elsewhere to match known digit preferences in humanly generated African census data.

Our new tests provide a particularly powerful toolkit for monitoring budget expenditures and uncovering suspected fraud. This method works even when on the ground monitoring is challenging, as is often the case in remote and insecure parts of the developing world. In addition, it requires minimal cooperation from those inside the organization or government who may have an incentive to impede an investigation. In developing countries, where one faces strong corruption cartels and weak rule of law with which to force compliance, this is a major benefit.

One of our new tests, employing the generalized Benford's Law to analyze multiple digit places, provides a statistically powerful test applicable to even relatively small datasets. The ability to work on smaller sample sizes allows more multi-dimensional analyses, such as our comparisons across districts, years, and sectors. By partitioning by project years, we are able to demonstrate that more suspicious patterns emerge in a presidential election year, consistent with reports that funds were illegally diverted to fund political campaigns.

Our new test of overuse of high digits in valuable digit places uncovers strategic and profitable deviation. This is the first test we know of that relates aberrant digit patterns to the monetary value of the digit place. Furthermore, unpacking rounded numbers reveals evidence of more fraudulent patterns than unrounded numbers. This might arise when data fabrication is the result of low-effort human generation and when officials fabricate data to achieve specific embezzlement targets. Both tests point to evidence of intent, which has had added significance for legal proceedings.

Readers may be concerned that publication of these methods will provide potential fraudsters with the means to beat the monitors. They need not worry. Engineering a Benford conforming dataset is a vastly more challenging statistical exercise than is ensuring that digits are uniformly distributed. It would also require centralization across the organization, and matching of all supporting documentation, such as coordination of date stamped receipts, cashbooks, vehicle logs, cancelled checks, and bank statements. Furthermore, each individual fabricating data would have an incentive to deviate to hide higher levels of theft, undercutting efforts to produce aggregate results consistent with Benford's Law. Such high levels of coordination would also place leadership at high risk of detection.

In the development of new methods, external validity is a paramount concern. Our methods are validated in two ways. First, the World Bank conducted a forensic audit of this same project, using field-verified, transaction-level auditing procedures different from the methods that we employ. Our method finds similar levels and locations of suspected fraud. Second, our method examines patterns in two types of data that were produced in different ways: expenditures that are sums of receipts, and participant data that are counts of people. We find the same aberrant digit patterns in both types of data, a strong indicator of human tampering.

The validity and expanded capabilities of our methods will facilitate broader use of digit analysis, especially where sample size is an issue, and data partitioning is desirable. The areas that might benefit, and where digit analysis has already been used, are in auditing, election fraud, scientific data fabrication, and the monitoring of enumerator integrity during survey research. For example, our method could have been used in the forensic audit of this World Bank project to identify and target the worst offending districts, two of which were missed in the World Bank audit sample. Such an application can provide enormous efficiency gains for auditing. We also foresee use in a variety of new applications, for example, to check the authenticity of data supplied by governments in compliance with international ecological and environmental agreements, pollution and labor data supplied for treaty compliance, and data supplied to the IMF in conjunction with lending. Our digit analysis methods provide an efficient, cost-effective, and scalable way to uncover suspected fraudulent patterns without requiring much cooperation from those who may wish to impede such investigations.

In recent years tens of billions of aid dollars have been squandered in mega-corruption scandals, including those in Iraq, Afghanistan, Syria, Somalia, and the health initiatives of the Global Fund. The new digit analysis methods that we present in this paper provide a means for monitoring expenditures even where oversight is difficult and dangerous, and cooperation from governments is not forthcoming. Our techniques also have innumerable applications for detecting fraud in many other forms of data, including: survey data, scientific data, and environmental data supplied by nations to demonstrate compliance with international agreements. As big data skyrockets in modern government and industry, it is increasingly important to have techniques for monitoring its authenticity and measuring the level of potential fraud.

CALIFORNIA INSTITUTE OF TECHNOLOGY

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

Alvarez, R. Michael, Thad E. Hall, and Susan D. Hyde, eds., *Election Fraud: Detecting and Deterring Electoral Manipulation* (Brookings Institution Press, 2008).

Amiram, Dan, Zahn Bozanic, and Ethan Rouen, "Financial statement errors: Evidence from the distributional properties of financial statement numbers," *Review of Accounting Studies*, 20 (2015), 1540-1593.

Avis, Eric, Claudio Ferraz, and Frederico Finan, "Do government audits reduce corruption? Estimating the impacts of exposing corrupt politicians," *National Bureau of Economic Research Working Paper Series*, No. 22443 (2016).

Banerjee, Abhijit, Sendhil Mullainathan, and Rema Hanna, "Title," National Bureau of Economic Research Working Paper No. 17968, 2012.

Beber, Bernd, and Alexandra Scacco, "What the numbers say: A digit-based test for election fraud," *Political Analysis*, 20 (2012), 211-234.

Becker, Gary S, "Crime and punishment: An economic approach," *Journal of Political Economy*, 76 (1968), 169-217.

Besley, Timothy, and John McLaren, "Taxes and Bribery: The Role of Wage Incentives," *The Economic Journal*, 103 (1993).

Boland, Philip, and Kevin Hutchinson, "Student selection of random digits," *The Statistician*, 49 (2000), 519-529.

Boyle, Jeff, "An application of Fourier series to the most significant digit problem," *The American Mathematical Monthly*, 101 (1994), 879-886.

Bredl, Sebastian, Peter Winker, and Kerstin Kötschau, "A statistical approach to detect interviewer falsification of survey data," *Survey Methodology*, 38 (2012), 1-10.

Burgess, Robin, Matthew Hansen, Benjamin A Olken, Peter Potapov, and Stefanie Sieber, "The political economy of deforestation in the tropics," *The Quarterly Journal of Economics*, 127 (2012), 1707-1754.

Chapanis, Alphonse, "Human production of "random" numbers," *Perceptual and Motor Skills*, 81 (1995), 1347-1363.

Chassang, Sylvain, and Gerard Padró i Miquel, "Title," National Bureau of Economic Research Working Paper No. 20315, 2014.

Cho, Wendy K Tam, and Brian J Gaines, "Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance," *The American Statistician*, 61 (2012), 218-223.

da Silva, Carlos Gomez, and Pedro M.R. Carreira, "Selecting audit samples using Benford's Law," *Auditing: A Journal of Practice and Theory*, 32 (2013), 53-65.

Deckert, Joseph, Mikhail Myagkov, and Peter Ordeshook, "Benford's Law and the detection of election fraud," *Political Analysis*, 19 (2011), 245-268.

Diekmann, Andreas, "Not the first digit! Using Benford's Law to detect fraudulent scientific data," *Journal of Applied Statistics*, 34 (2007), 321-329.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India," *The Quarterly Journal of Economics*, 128 (2013), 1499-1545.

Durtschi, Cindy, William Hillison, and Carl Pacini, "The effective use of Benford's Law to assist in detecting fraud in accounting data," *Journal of Forensic Accounting*, V (2004), 17-34.

Ferraz, Claudio, and Frederico Finan, "Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes," *The Quarterly Journal of Economics*, 123 (2008), 703-745.

Fisman, David, Ray Fisman, Julia Galef, Rakesh Khurana, and Yongxiang Wang, "Estimating the value of connections to Vice-President Cheney," (Working paper, Columbia University, 2006).

Fisman, Raymond, "Estimating the value of political connections," *The American Economic Review*, 91 (2001), 1095-1102.

Hill, Theodore P., "A statistical derivation of the significant-digit law," *Statistical Science*, 10 (1995), 354-363.

Lindijer, Koert, "Kenya is run by mafia-style cartels, says chief justice," in *The Guardian*, (2016).

Lui, Francis T., "A Dynamic Model of Corruption Deterrence," *Journal of Public Economics*, 31 (1986), 215-236.

Mcmillan, John, and Pablo Zoido, "How to subvert democracy: Montesinos in Peru," *Journal of Economic Perspectives*, 18 (2004), 69-92.

Michalski, Tomasz, and Gilles Stoltz, "Do Countries Falsify Economic Data Strategically? Some Evidence That They Might," *The Review of Economics and Statistics*, 95 (2013), 591-616.

Nagi, M.H, E.G. Stockwell, and L.M. Snavley, "Digit preference and avoidance in the age statistics of some recent African censuses: Some patterns and correlates," *International Statistical Review*, 41 (1973), 165-174.

Nigrini, M, *Benford's Law* (Hoboken, New Jersey: John Wiley & Sons, Inc, 2012).

Nigrini, M, and L Mittermaier, "The use of Benford's Law as an aid in analytic procedures," *Auditing: A Journal of Practice and Theory*, 16 (1997).

Olken, Benjamin A, "Monitoring corruption: Evidence from a field experiment in Indonesia," *Journal of Political Economy*, 115 (2007), 200-249.

Olken, Benjamin A, and Patrick Barron, "The simple economics of extortion: Evidence from trucking in Aceh," *Journal of Political Economy*, 117 (2009), 417-452.

Rath, Gustave J., "Randomization by humans," *The American Journal of Psychology*, 79 (1966), 97-103.

Reinikka, Ritva, and Jakob Svensson, "Using micro-surveys to measure and explain corruption," *World Development*, 34 (2006), 359-370.

Republic of Kenya, " Arid Lands Resource Management Project (Phase II) Tana River

District Progress Report 2003-2006," (2006).

Schleifer, A, and R Vishny, "Corruption," *Quarterly Journal of Economics*, 108 (1993), 599-617.

Schräpler, Jörg-Peter, "Benford's Law as an instrument for fraud detection in surveys using the data of the Socio-Economic Panel (SOEP)," *Journal of Economics and Statistics*, 231 (2011), 685-718.

UN Economic and Social Council Economic Comission for Africa, "Adjustment of errors in the reported age-sex data from African censuses," in *Joint Conference of African Planners, Statisticians and Demographers*, (Addis Ababa, Ethiopia, 1986).

Walter R. Mebane, Jr., "Comment on "Benford's Law and the detection of election fraud"," *Political Analysis*, 19 (2011), 269-272.

World Bank, "Project appraisal document on a proposed credit in the amount of SDR 43.6 Million (US$ 60M. equivalent) to the Republic of Kenya for the Arid Lands Resource Management Project Phase Two," Country Department 5 Eastern and Southern African Rural Development Operations, Africa Region, ed. (2003).

---, "Forensic audit report: Arid Lands Resource Management Project - Phase II," (Redacted: World Bank, 2011).

TABLE 1: EXPECTED DIGIT FREQUENCIES UNDER BENFORD'S LAW

		Digit Place				
		1	2	3	4	5
	0	0.0000	0.1197	0.1018	0.1002	0.10002
	1	0.3010	0.1139	0.1014	0.1001	0.10001
	2	0.1761	0.1088	0.1010	0.1001	0.10001
	3	0.1249	0.1043	0.1006	0.1001	0.10001
Digit	4	0.0969	0.1003	0.1002	0.1000	0.10000
	5	0.0792	0.0967	0.0998	0.1000	0.10000
	6	0.0669	0.0934	0.0994	0.0999	0.09999
	7	0.0580	0.0904	0.0990	0.0999	0.09999
	8	0.0512	0.0876	0.0986	0.0999	0.09999
	9	0.0458	0.0850	0.0983	0.0998	0.09998

Notes: Source is (Nigrini and Mittermaier 1997, pp. 54).

TABLE 2. SIGNIFICANCE OF DIGIT TESTS BY DISTRICT

Fig	Digit Test	Baringo	Garissa	Ijara	Isiolo	Mandera	Marsabit	Moyale	Samburu	Tana	Turkana	Wajir	All Districts
1	First Digit: Expenditure Data	**1.4E-09** **488**	0.029 430	**2.3E-13** **386**	**5.5E-06** **308**	**1.4E-08** **489**	0.011 293	**1.9E-12** **319**	**5.7E-05** **359**	**0.0037** **332**	0.071 357	0.37 578	0.089 4339
2	Second Digit High: Expenditure	**1.4E-04** **397**	0.38 331	0.073 287	0.13 182	0.034 367	0.26 211	**0.0022** **239**	0.50 275	0.094 250	0.16 262	0.012 422	**1.1E-04** **3223**
3	Last Digit: Participant	**2.0E-09** **635**	**0.0015** **428**	**9.9E-06** **685**	**2.3E-08** **443**	**1.7E-10** **711**	0.018 428	**8.2E-04** **704**	**7.2E-05** **575**	0.17 474	0.95 404	**0.0035** **363**	**7.0E-26** **5850**
4A	Rounding Digits: Expenditure				**Top Quartile**	**Top Quartile**	**Top Quartile**						DNA
4B	Repeating Numbers: Expenditure	**Top Quartile**			**Top Quartile**	**Top Quartile**							DNA
5	Sector Effects: Expenditure	**> 3x**		**> 3x**	**> 3x**	**> 3x**			**> 3x**			**>3x**	DNA
6	Digit Pairs: Participant	**8.2E-04** **251**	**1.8E-04** **293**	**0.0037** **176**	**0.0044** **125**	0.0091 238	**0.0031** **126**	0.41 173	0.38 166	0.035 137	0.49 119	**7.1E-05** **255**	**1.4E-09** **2059**
7A	All Digit Places beyond the First: Expenditure	**7.2E-17** **1352**	**2.8E-08** **976**	**2.6E-05** **769**	0.0082 437	**3.6E-14** **846**	**3.9E-04** **449**	**1.5E-14** **671**	0.020 848	**7.8E-04** **868**	0.40 907	**1.9E-06** **1248**	**3.9E-15** **9371**
7B	All Digit Places beyond the First: Participant	**2.1E-04** **674**	**6.2E-18** **858**	**5.9E-11** **765**	**6.1E-11** **478**	**9.0E-18** **886**	0.25 527	0.033 736	**2.3E-05** **639**	0.013 500	**0.0037** **591**	**6.5E-15** **731**	**5.7E-51** **7385**
8	Year Effects (2007): Expenditure	0.0073 182	0.016 139	0.045 222	0.22 98	0.18 117	0.032 88	0.033 238	0.15 231	0.75 165	**0.0045** **192**	0.088 273	**6.5E-06** **1945**
9	Unpacking Rounded Numbers: Participant	0.0085 248	**5.9E-24** **459**	**1.2E-10** **298**	**4.4E-13** **157**	**6.1E-21** **453**	0.014 222	**0.0030** **179**	**3.9E-05** **179**	0.057 142	**3.1E-05** **205**	**7.6E-11** **433**	**2.6E-64** **2975**
10	Deviation from Mean in 10,000s Digit Place	0.024	**0.0015**	0.0054	0.131	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**
	Number of Significant Tests $p < 0.005$ (Out of 12)	8	6	7	8	9	4	6	6	3	4	7	9

Notes: We ran 12 digit tests on each of 11 districts. The tests were chosen to reduce overlap in measurement. Given the large number of tests, a Bonferroni correction was used to establish 0.005 as the acceptable p – value for 9 of our tests. Failed tests at the 0.005 level are indicated in bold. For rounding and repeating (Figures 4A and 4B), there is no theoretical means to establish the expected level and we work from the null hypothesis that there should be no significant difference between the districts. We flag the districts that are outliers in the upper quartile. Similarly, for the sector analysis (Figure 5), we flag the districts for which repeated numbers in sectors with higher risk of fraud (training and transport) are more than triple the level of other sectors (civil works and goods and equipment). We tabulate the number of significant tests for each district in the bottom row.

TABLE 3. DIGIT TESTS BY DISTRICT COMPARED TO WORLD BANK INT FORENSIC AUDIT RESULTS

	Number of Failed Digit Tests out of 12 ($p < 0.005$)	INT Audit (Percent Suspected Fraudulent and Questionable Transactions)
Isiolo	8	74
Wajir	7	75
Samburu	6	68
Garissa	6	62
Tana	3	44
Mandera	9	Not Audited
Baringo	8	Not Audited
Ijara	7	Not Audited
Moyale	6	Not Audited
Marsabit	4	Not Audited
Turkana	4	Not Audited

Note: Source for the INT forensic audit data is (World Bank 2011).

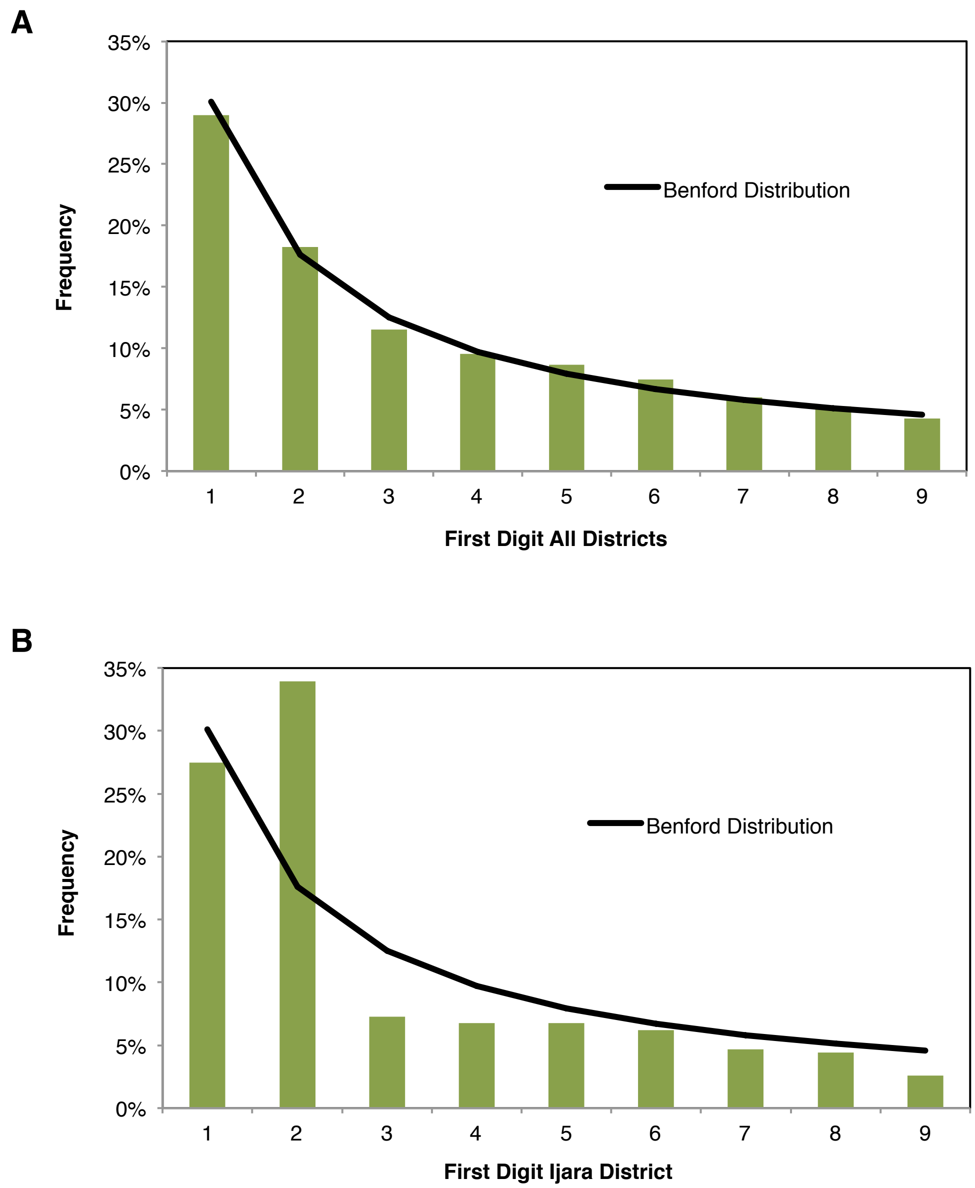


FIGURE 1. FIRST DIGIT EXPENDITURE DATA AGAINST BENFORD'S LAW

Notes. (A) All districts combined (p = 0.089; n = 4339). (B) Ijara District only ($p = 2.3 \times 10^{-13}$; n = 386).

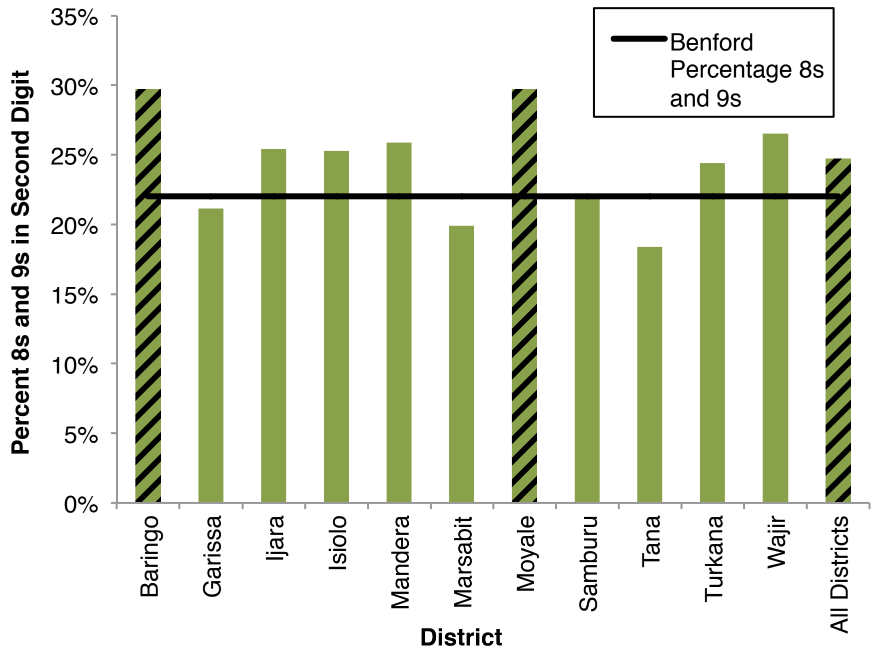


FIGURE 2. HIGH SECOND DIGITS IN EXPENDITURE DATA BY DISTRICT

Notes: Figure 2 tests the observed frequency of 8s and 9s in the second digit place against the expected frequency of 22 percent given by Benford's Law. Significance is tested in a binomial test and shown with crosshatching. The results are significant for two districts (Baringo $p = 1.4 \times 10^{-4}$, $n = 397$ and Moyale $p = 0.0022$, $n = 239$) and for all districts ($p = 1.1 \times 10^{-4}$, $n = 3223$).

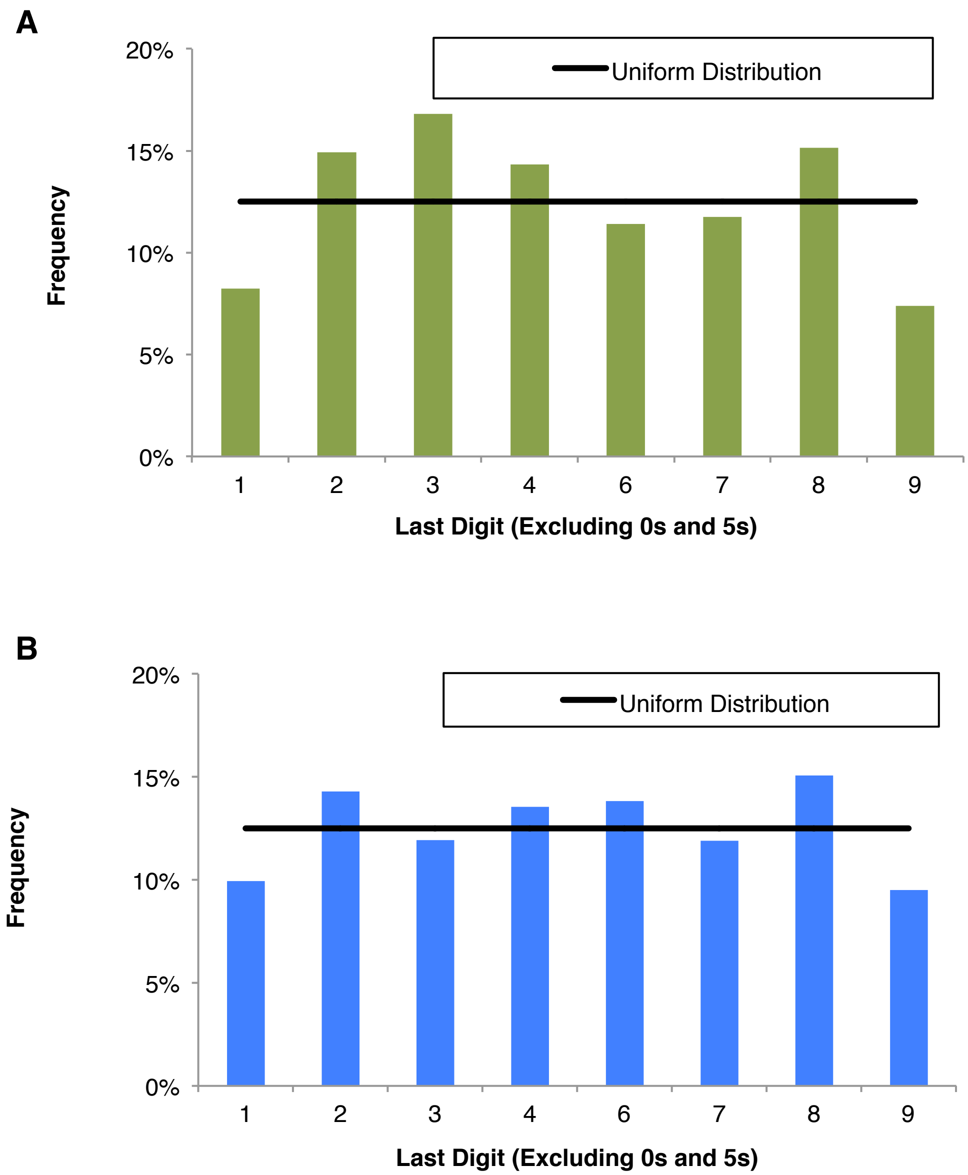


FIGURE 3. LAST DIGIT EXPENDITURE AND PARTICIPANT DATA AGAINST THE UNIFORM DISTRIBUTION.

Notes. (A) Expenditure data ($p = 1.5 \times 10^{-9}$; $n = 851$). (B) Participant data ($p = 7.0 \times 10^{-26}$; $n = 5850$).

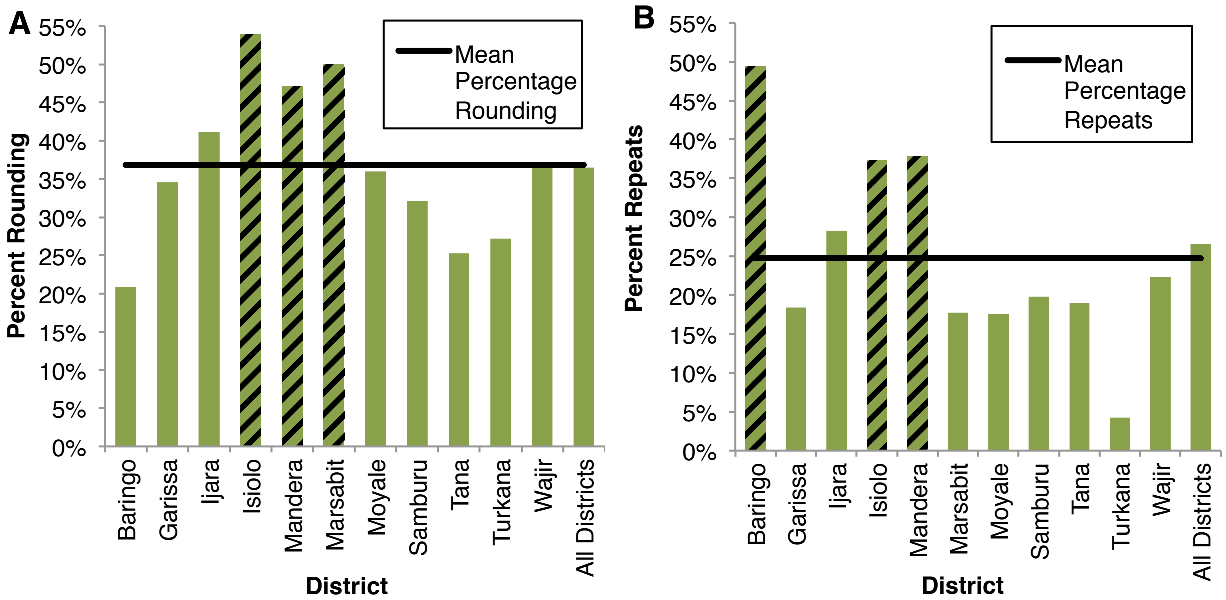


FIGURE 4. PERCENTAGE OF ROUNDED DIGITS AND REPEATED NUMBERS IN EXPENDITURE DATA BY DISTRICT

Notes. (A) Percentage digit places rounded in expenditure data by district. (B) Percentage repeated expenditure entries by district for a given annual report. The three districts representing the top quartile are crosshatched.

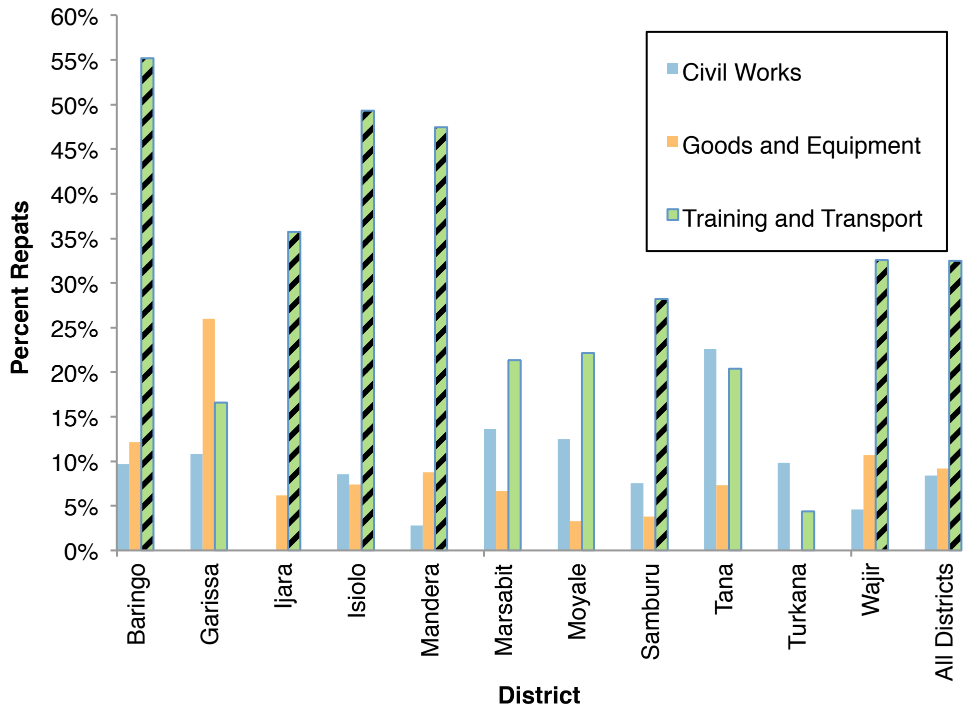


FIGURE 5. SECTOR EFFECTS IN EXPENDITURES

Notes. Percentage of line item expenditures repeated exactly by district, year, and sector. Crosshatched districts report over three times as many repeated numbers in training and transport, versus other sectors, consistent with low-effort data manipulation.

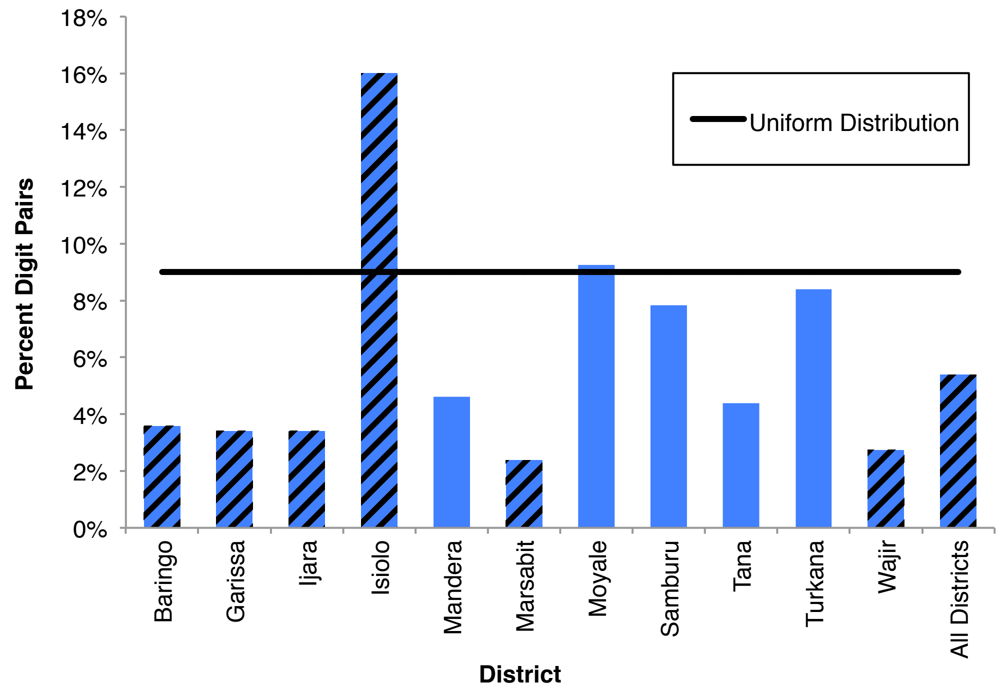
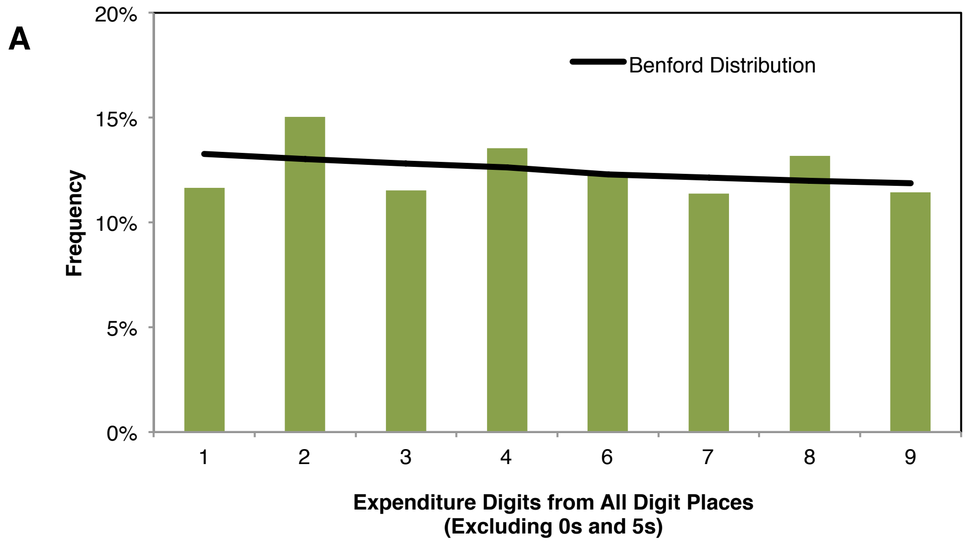


FIGURE 6. DIGITS PAIRS IN THE LAST TWO DIGITS FOR PARTICIPANT DATA BY DISTRICT

Notes. Crosshatched districts are statistically significantly failing the binomial test at $p < 0.005$ by over or underutilizing digit pairs such as 11, 22, and 33 (Baringo $p=8.2 \times 10^{-4}$, $n = 251$; Garissa $p = 1.8 \times 10^{-4}$, $n = 293$; Ijara $p = 0.0037$, $n = 176$; Isiolo $p = 0.0044$, $n = 125$; Marsabit $p = 0.0031$, $n = 126$; Wajir $p = 7.1 \times 10^{-5}$, $n = 255$; all districts $p = 1.4 \times 10^{-9}$, $n = 2059$).



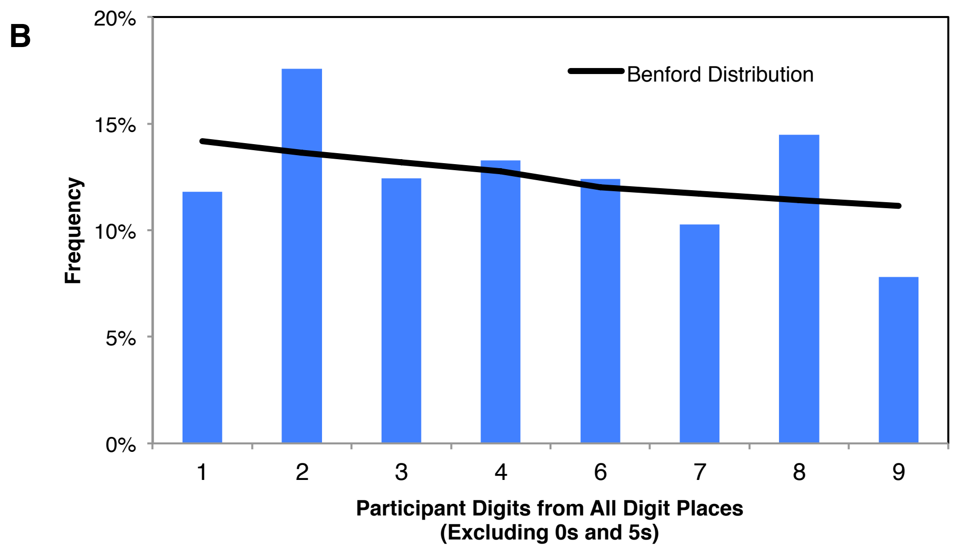


FIGURE 7. ALL DIGIT PLACES BEYOND THE FIRST AGAINST BENFORD'S LAW FOR EXPENDITURE AND PARTICIPANT DATA

Notes. (A) Expenditure data ($p = 3.9 \times 10^{-15}$; $n = 9371$). (B) Participant data ($p = 5.7 \times 10^{-51}$; $n = 7385$).

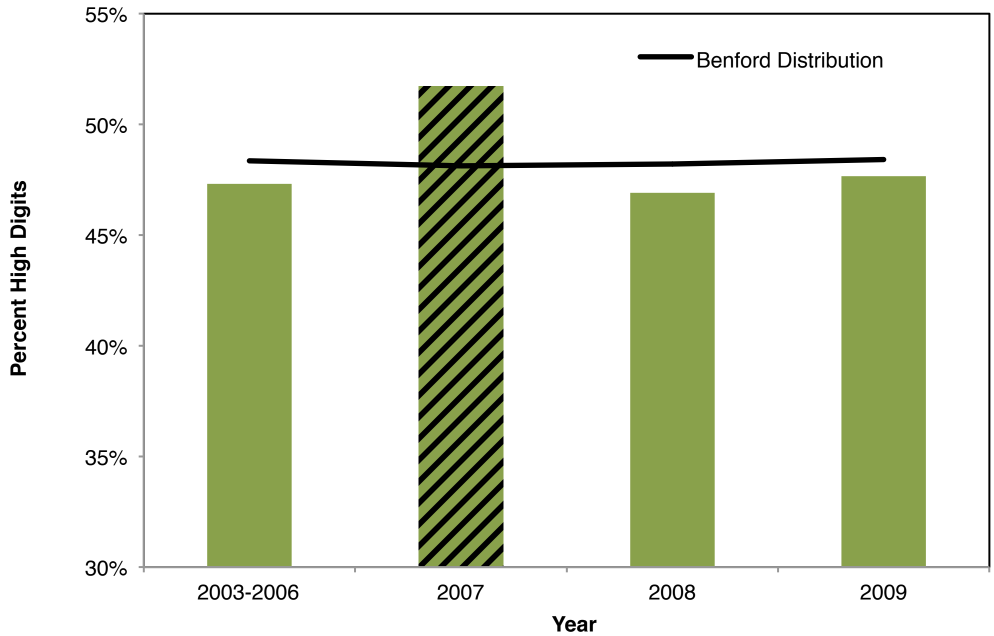
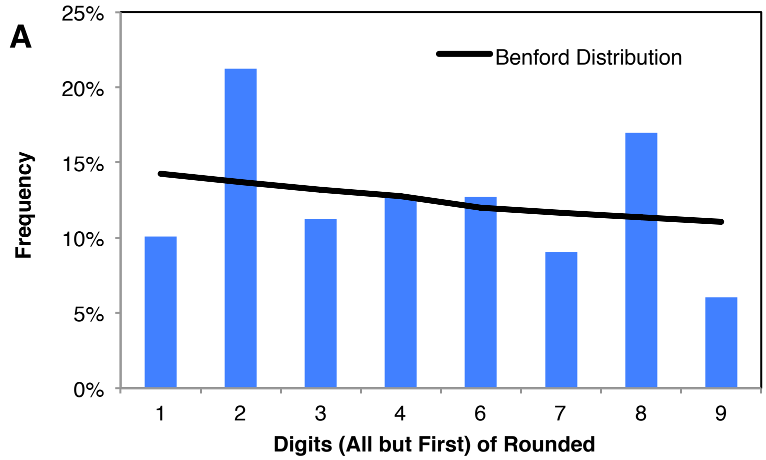


FIGURE 8. ELECTION YEAR EFFECTS IN EXPENDITURE DATA

Notes. Percentage of high digits (6, 7, 8, and 9) in all digit places but the first, for all districts, by year. 2007 was a Presidential election year. 2007 has a statistically significant presence of high digits in a ($p = 6.5 \times 10^{-6}$; $n = 1945$.)



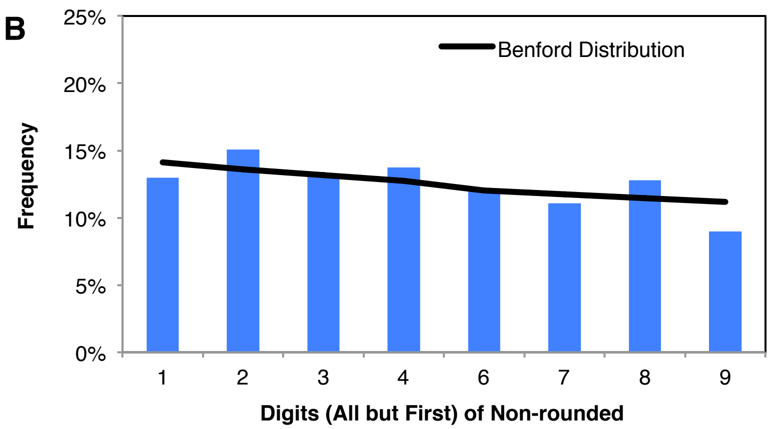


FIGURE 9. UNPACKING ROUNDED AND UNROUNDED DIGITS IN PARTICIPANT DATA

Notes: **(A)** Digit breakout of all but the first digit (excluding 0s and 5s) when the total of male and female participants sums to a rounded number ($p = 2.6 \times 10^{-64}$; $n = 2975$). **(B)** Digit breakout of all but the first digit (excluding 0s and 5s) when the total of male and female participants sums to a non-rounded number ($p = 1.9 \times 10^{-6}$; $n = 4410$).

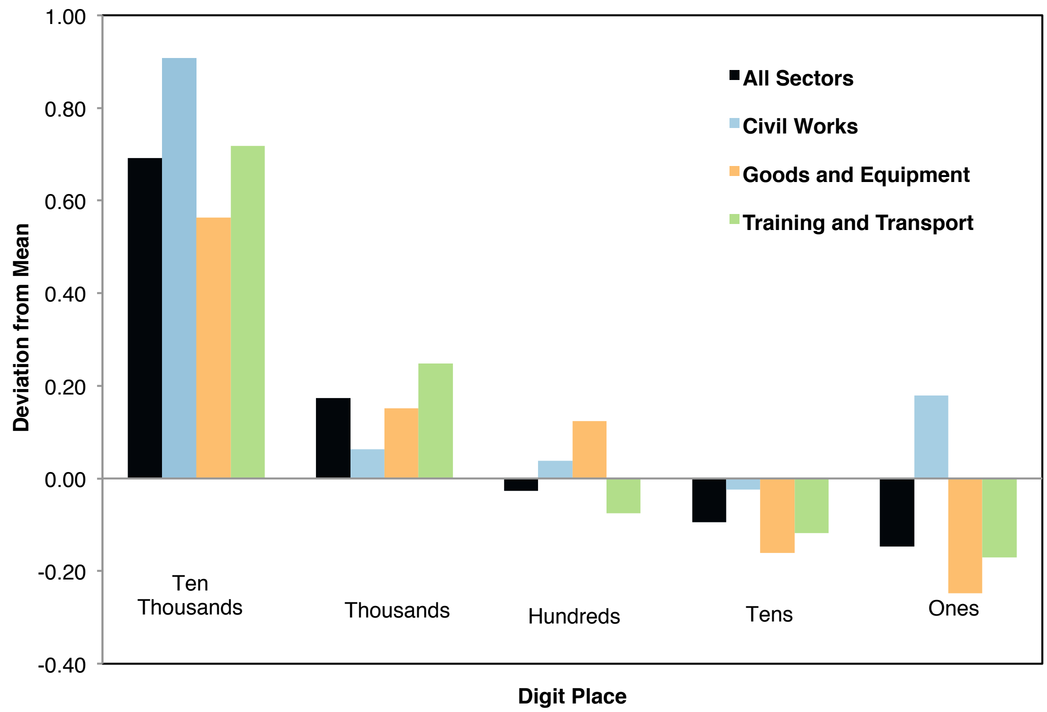


FIGURE 10. DEVIATION FROM BENFORD'S LAW MEAN IN EXPENDITURE DATA WITH MONTE CARLO SIMULATION

Notes. We compare the observed mean by digit place from the right to the Benford expected mean in each sector. Zero reflects conformance to the Benford expected mean. Positive values indicate the mean is higher than Benford's Law predicts. The observed pattern is consistent with a strategy of high digits in high digit value places and then underusing them in low digit value places to even out the digit distribution. We perform a Monte Carlo simulation of Benford-conforming datasets and compare our observed statistics to the simulated statistics to produce p-values. Compared to a sample of 100,000 simulations, using data from all sectors, we observe the following statistics: 10,000s place ($p = 1.0 \times 10^{-5}$), 1,000s ($p = 2.3 \times 10^{-4}$), 100s ($p = 0.33$), 10s ($p = 0.10$), 1s ($p = 0.061$).