World Development 188 (2025) 106858

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Regular Research Article

Detecting Corruption: Evidence from a World Bank project in Kenya

Jean Ensminger [a], Jetson Leder-Luis [b,*]

[a] California Institute of Technology, HSS MC 228-77, Pasadena, CA 91125, United States
[b] Boston University and NBER, 595 Commonwealth Avenue #518D, Boston, MA 02115, United States

ARTICLE INFO

JEL Codes:
D73
H83
O22
M42
C49

Keywords:
Development aid
Fraud
Detection

ABSTRACT

Corruption is a major problem in development aid, in part because areas with the greatest need for development assistance often have weak governance. In these environments, traditional anti-fraud measures such as audits or criminal enforcement are limited in their effectiveness. Moreover, aid organizations face incentives to downplay bad outcomes for fear of alienating donors, which has led to the suppression of negative findings related to development aid fraud.

In this paper, we develop new statistical tests to uncover strategic data manipulation consistent with fraud, which can help identify falsified data and facilitate monitoring in difficult-to-audit circumstances. We apply this method to a World Bank community driven development project in Kenya. Our statistical tests rely on the fact that human-produced digits and naturally occurring digits have different digit patterns: unmanipulated digits follow the Benford's Law distribution. We improve upon existing digit analysis techniques by being sensitive to the value of digits reported, which helps distinguish between intent to defraud and error, and by improving statistical power to allow for finer partitioning of the data. We also produce simulations that demonstrate the superiority of our new tests to the standards in the field, and we provide a new *R* package for conducting our statistical tests.

Our study finds substantial evidence of fraud, validated by qualitative data, a forensic audit conducted by the World Bank, and replication with a separate dataset for external validity. We uncover higher levels of fraud in a Kenyan election year when graft also had political value and in harder to monitor sectors. This methodology also has broad applications to many forms of data beyond those encountered in development aid.

1. Introduction

Fraud and corruption are major issues in the developing world. Developing countries, and the aid organizations that serve them, often operate in weak institutional environments where there are high opportunities for theft of resources. The primary mechanisms for detecting and deterring corruption and fraud —such as auditing, transparency, and criminal and civil liability for corrupt individuals—require strong institutions and accountability when rules or norms are violated. Therefore, these tools are most challenging to implement where they are most necessary, in governments with systemic corruption (Svensson, 2005). Moreover, aid organizations that serve developing countries face these challenges on the ground, but also have strong incentives not to report their own failures, for fear of losing the support of donors. These agency issues have hindered the application of traditional anti-fraud policy in the development aid space.

In this paper, we provide new methods for detecting fraud and apply those methods to reveal important substantive findings about fraud in a large World Bank development project. Our tools are based on digit analysis, which analyzes the patterns of reported data to detect fraud. These tools rely on the fact that humanly generated data are different from naturally occurring data. Humans face incentives to manipulate data, as well as behavioral biases when producing data, while naturally occurring data follow Benford's Law. We build upon earlier digit analysis work to improve statistical power and present new tests that better reveal suspected intent to defraud.

We apply our method to data from a World Bank development project in Kenya. This project was chosen for study because it has two advantages over alternative projects. First, it is a Community Driven Development (CDD) project, which represents a common development strategy of the World Bank that has been adopted worldwide among many donors. Second, the project has the unique advantage that it was ultimately subject to a forensic audit by the World Bank. This allows us to verify our digit analysis results against a forensic audit of the same

* Corresponding author.
E-mail addresses: jensming@caltech.edu (J. Ensminger), jetson@bu.edu (J. Leder-Luis).

https://doi.org/10.1016/j.worlddev.2024.106858
Accepted 13 November 2024
Available online 7 December 2024

project.

Our results confirm the high levels of graft documented in qualitative work on the project (Ensminger, 2017) and in interview data, which we present in Appendix A. The analysis also reveals some important substantive findings that underscore our methods' advantages. First, we find significant inflation of expenditures during the 2007 Kenyan presidential election year. This is consistent with our qualitative data that World Bank funds were syphoned into the Kenyan presidential election campaign of 2007, which is widely accepted to have been a stolen election (Gibson & Long, 2009). Second, our tests reveal higher levels of manipulation in harder-to-monitor types of spending, consistent with a rational crime approach (Becker, 1968) and previous empirical results (Karpoff, Lee, & Vendrzyk, 1999) (Dávid-Barrett & Fazekas, 2020).

Our statistical tests analyze different patterns of reported data and find behavior consistent with fraud. We perform these tests on both the line-item expenditures reported in the project as well as the reported counts of beneficiaries served by the project. Naturally occurring data and humanly produced data are different both because humans face behavioral limitations in producing numbers (Chapanis, 1995), and individuals have incentives to pad values in response to their economic and political environments. In contrast, naturally occurring data follow Benford's Law, a logarithmic distribution that gives probabilities of digits in each digit place, where low digits (1, 2, etc.) are more likely to appear closer to the front of a number. Our work sharpens this technique. First, we expand the statistical power of Benford's Law goodness of fit testing by considering all digit places in one test. Second, we build a new type of digit analysis that considers the *value* of the number, which allows us to distinguish between patterns consistent with profitable misreporting and those created by benign errors. We then supplement our 2 new tests with the results of 8 other tests, including 2 applications of these new methods and 6 tests from the existing literature, for 10 tests in total.

We validate the results of our analysis using a unique complementary data source: the forensic audit of the same World Bank project. In response to an external complaint, the World Bank conducted a two-year forensic audit of the project (World Bank Integrity Vice Presidency, 2011). The audit revealed that the Bank's financial controls, monitoring, and existing audit mechanisms were not capturing the extreme level of suspected fraud. The World Bank forensic audit flagged 66 % of district transactions as suspicious (49 % as suspected fraudulent and 17 % as questionable). [1]The findings of this audit validate our statistical methodology: the number of tests that show statistically significant deviations per geographic district is correlated ($p < 0.05$) with the level of suspected fraudulent and questionable transactions from the same districts examined in the forensic audit.

We conduct two further exercises to validate our statistical tests. First, we run simulations to show that our tests can successfully detect misreporting in a way that is not specific to this case study. Our simulations also test alternative hypotheses, showing for example that the patterns we uncover are not driven by benign factors such as underlying prices. Second, to externally validate our methods, we repeat our digit analysis on another dataset of self-reported revenues and employee counts from the World Bank Enterprise Surveys. We show that the level of behavioral biases in reported data is statistically significantly negatively correlated with the Control of Corruption measure from the World Governance Indicators.

Our method is broadly applicable and can be useful for detecting fraud in a variety of contexts. Because our statistical tests apply both to financial data and also quantities, we foresee its usefulness for other areas of concern in developing countries, such as over-invoicing on construction projects, data submission to meet environmental and loan compliance requirements, and pharmaceutical theft (Transparency International, 2016).

1.1. Literature review

A large body of literature has addressed fraud and corruption in the context of development. Olken and Pande (2012) provide an overview of the major topics surrounding developing world corruption, including magnitudes, efficiency costs, determinants of corruption, and effective policy solutions. One major theme in this literature is the effectiveness of audits. Olken (2007) uses experiments in Indonesia to show that monitoring is effective at reducing fraud in infrastructure expenses. Ferraz and Finan (2008) show that random city audits in Brazil expose corruption and impact incumbents' electoral performance. Duflo et al. (2013) provide an example of auditor capture in India and show that monitoring of monitors is an effective way to combat fraud. However, Cuneo et al. (2023) note that the results of these studies may be limited to middle-state capacity environments and they discuss why audits are ineffective in very low state capacity environments, which corroborates our evidence that auditing was ineffective in the Kenyan context. Dávid-Barrett and Fazekas (2020) show that anti-fraud efforts in World Bank procurement can lead to fraud being diverted, rather than eliminated, to evade detection. Little work in development economics has addressed detecting fraud, per se.

A more limited body of work has studied the incentives of development aid organizations. Lamoreaux et al. (2015) find that accounting issues in World Bank development aid loans are more likely to be overlooked in areas of strategic importance for U.S. interests. Andersen et al. (2022) provide evidence of offshoring of World Bank funds; we discuss this paper, and World Bank attempts to suppress it, in the next section. These papers establish the need for new tools to address fraud in challenging environments.

Digit analysis and Benford's Law have generated a long literature of statistical methods. Digit analysis has been used in corporate accounting to measure financial statement errors (Amiram, Bozanic, & Rouen, 2015), as well as in forensic auditing, where it is used for targeting deeper investigations (Nigrini & Mittermaier, 1997; Durtschi, Hillison, & Pacini, 2004). Digit analysis has also had widespread application to other areas where there is value in detecting data manipulation. Digit analysis has been used extensively in the detection of election fraud (Mebane, 2008; Beber & Scacco, 2012; Mack & Stoetzer, 2019), as well as in the detection of IMF data manipulation (Michalski & Stoltz, 2013), campaign finance fraud (Cho & Gaines, 2012), scientific data fabrication (Diekmann, 2007) (Toedter, 2019), and enumerator integrity during survey research (Bredl, Winker, & Kötschau, 2012; Judge & Schechter, 2009; Schräpler, 2011).

Recent advances in Benford's Law testing have improved statistical precision and power. Nigrini and Miller (2009) employ a second-order test of conformance to Benford's Law, which considers the difference between ranked values in a dataset. Da Silva and Carreira (2013) use Benford's law to find specific subsets of the data with the greatest nonconformance that can be used to guide audits. Barabesi et al. (2018) apply digit analysis tests to detect customs fraud using a sequential testing procedure, testing multiple high-level hypotheses and then lower-level single-digit hypotheses. Cerioli et al. (2019) apply a different method to international trade data, using corrected test statistics that account for false positives, given that values in international trade data may not conform to Benford's Law. In each of these papers, the authors conduct tests for conformance to the Benford distribution to improve power or target their test or their sample based solely on the Benford's Law distribution.

2. Background: Auditing and development aid

From 2010 to 2020, aid to developing countries totaled \$1.7 trillion (OECD, 2022). Developed nations around the world make sizeable

[1] Diversion of funds to this degree has been reported in related contexts as well. Reinikka and Svensson (2004) find 87% diversion of funds in a Ugandan government education grant program.

investments in projects to promote growth and development in poor countries. Ensuring that these funds are spent appropriately is critical to the effectiveness of development aid.

The empirical relationship between aid and corruption highlights the need for detection mechanisms, because aid flows disproportionately to nations with weak institutions. Figure 1 shows the correlation between net aid flows and the Worldwide Governance Indicator measure of the perception of corruption levels by country in 2019 (The World Bank, 2019) (Kaufmann & Kraay, 2020). This figure makes two points. The slope of the linear regression between log aid dollars and corruption control is −0.95, (p = 0.000, 95 % confidence interval [-1.3, −0.6]), indicating a statistically significant correlation. Moreover, of the $115 Billion in foreign aid to countries in these data, 92 % of aid flows to countries with a below-mean corruption control measure, indicating the scale of the threat that aid dollars face.

The World Bank has historically relied on internal investigations and monitoring tools such as routine financial review, supervisory missions, internal audits, and whistleblower hotlines as its primary anti-fraud mechanisms (Aguilar, Gill, & Pino, 2000). Gans-Morse et al (2018) compare strategies to reduce bureaucratic corruption, finding monitoring most effective. Easterly and Williamson (2011) faults nearly all agencies for poor transparency. Similarly, Alt (2018) highlights the importance of government budget and financial transparency in lowering corruption. Overall, the usefulness of these tools relies on the ability and willingness of development aid staff to make internal reports, conduct investigations, disseminate those findings, and take corrective action. Management must also make sufficient funds and staffing available to ensure adequate monitoring.

From a practical standpoint, there are many reasons why audits in developing contexts are challenging. Development aid projects span a variety of sectors, and include infrastructure, goods and equipment, services such as health care or child education, and trainings for beneficiaries to improve their human capital in areas such as agriculture. These projects, which generally reimburse costs, face serious monitoring challenges. Infrastructure projects, such as the construction of a school or a well, can face issues with low quality material or over-invoicing. Auditing the quality of materials is challenging and may necessitate a highly trained surveyor (Olken, 2007) This is particularly difficult when the projects occur in rural, dangerous, and hard to access parts of developing countries. Trainings and services produce even less physical evidence, and auditors may not be able to find beneficiaries to confirm expenditures. Beneficiaries may also face retaliation from the project for giving negative statements to outside monitors.

Development organizations also face conflicts of interest. These organizations often depend upon the field-supervision of outside experts who are typically chosen by the staff members overseeing the project. Their employment on future projects may depend upon favorable reports. Routine financial management and auditing is usually handled internally by understaffed departments.

When internal monitoring or external complaints at the World Bank point to potential fraud, the World Bank Integrity Vice Presidency (INT) is responsible for the Bank's fraud investigations; similar responsibilities are held by the Office of the Inspector General for USAID (OIG-USAID) and the European Anti-Fraud Office (OLAF) for the EU. In Fiscal Year 2021, World Bank INT received 4,311 complaints, but opened only 347 investigations, and produced only 35 sanctions or settlements (World Bank Group Sanctions System, 2021). Similarly, in Fiscal Year 2021, the OIG-USAID reported $4.9 billion in audited funds out of its $19.6 billion budget, with only 142 investigations closed (U.S. Agency for International Development Office of Inspector General., 2021a; U.S. Agency for International Development Office of Inspector General., 2021b). This paper uses data from a rare forensic audit of the World Bank. According to the then head of anti-corruption investigations at the World Bank (Stefanovic, 2018), no other field-verified, transaction-level, forensic audit of this scope had taken place for any World Bank project before or since the one we study.

A primary factor in the low rates of auditing in developing contexts is the lack of incentives to monitor and the incentives not to disclose

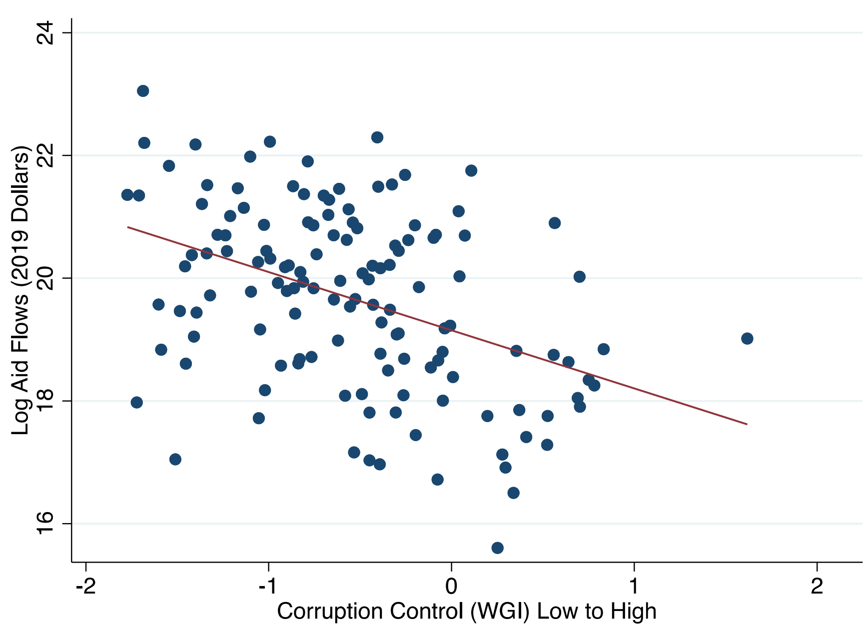

Figure 1. Corruption control and aid flows.
Notes: This figure plots the Worldwide Governance Indicator (WGI) control of corruption measure against log aid flows in 2019. WGI control of corruption measures "perceptions of the extent to which public power is exercised for private gain," standardized to mean 0 and standard deviation 1 (Kaufmann & Kraay, 2020); lower values correspond to lower controls and more corruption. Countries with worse corruption controls receive more aid. The slope of the linear regression is −0.95, (p = 0.000, 95 % confidence interval [-1.3, −0.6]). Log net aid flows are taken from the World Bank net official development assistance and official aid received and are measured in 2019 US dollars (World Bank and OECD Development Assistance Committee, 2023).

negative findings. The World Bank and other development organizations rely on funding from developed nations; in the U.S., aid is appropriated by Congress. Congress therefore faces a classic principal-agent problem under information asymmetry, as they are unable to properly monitor the effectiveness of these aid organizations. Aid is in this way a credence good (Dulleck & Kerschbamer, 2006): the principals, developed countries, must rely on the agents, the development aid organizations, both to administer the aid and to monitor their own performance. When development aid organizations uncover waste, fraud, or abuse, they stand to lose the support of donors, and therefore face strong incentives to hide the results of their findings, or not find fraud in the first place. This relates to the more general critique that aid agencies define their output in terms of money disbursed rather than services delivered (Easterly, 2002).

A recent research controversy demonstrates the incentives of development aid organizations to suppress evidence of corruption. In scholarly research, Andersen, Johannesen and Rijkers (2022) showed that aid disbursements to countries correspond to increases in deposits in offshore financial havens known for secrecy, amounting to 5–7.5 % of aid flows. One of the authors of that study, Bob Rijkers, is an employee of the World Bank, and his attempts to publish that piece were initially blocked by World Bank officials. World Bank employees and consultants are contractually bound to receive approval prior to publishing. In this controversial case, the Bank's Chief Economist resigned unexpectedly in protest, shortly following this incident (Jones, 2020) (The Economist, 2020). This case underscores the missing incentives for development aid organizations to effectively monitor themselves and disclose negative findings. Similar issues have been addressed qualitatively by Jansen (2013), who discusses the lack of oversight and incentives not to disclose negative findings in a natural resource management program in Tanzania funded by the Norwegian government. Jansen, who served as the program officer, notes the lack of external monitoring and attempts to suppress internal monitoring that are consistent with these misaligned incentives.

This paper proposes a partial solution to these challenges of monitoring, auditing, and misaligned incentives: the use of digit analysis to monitor development aid expenditures. Digit analysis requires development aid organizations to release project data that they already collect. In the interest of transparency, and to encourage independent external monitoring, this disclosure could be mandated by donor nations as a condition of funding bilateral and multi-lateral aid. Digit analysis also does not require the cooperation of potentially complicit subjects and can be used to detect early signs of fraud and to guide deeper investigations. This should help those responsible for project oversight within development organizations. By mandating data transparency, rather than pushing aid organizations to audit, donors can more easily ensure compliance. Digit analysis could then also be conducted by third parties, such as in-country beneficiaries, academics, anti-corruption organizations, and donor governments, who do not face the same conflicts of interest as those within the aid organizations.

3. Statistical theory

We motivate our statistical testing with a theoretical framework for the incentives of those who are tasked with producing expenditure reports. Those who produce reports, typically bureaucrats, face a decision either to accurately record spending or to fabricate such data. The statistical properties of the observed data result from this decision, and this theoretical framework provides predictions for the differences between legitimate and fabricated data.

3.1. *The statistical properties of truthfully reported data*

Using a set of receipts dedicated to a single transaction, such as the construction of a classroom, an honest bureaucrat calculates the sum of all the construction related receipts and enters the total in the report. These data follow the digit patterns of natural data.

Benford's Law describes the natural distribution of digits in financial data. Benford's Law is given mathematically by (Hill, 1995):

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

We have, for example, the probability that the first 3 digits are "452":

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

In the first digit place, Benford's Law 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 & Mittermaier, 1997). Table 1 shows the full digit-by-digit-place table of expected frequencies under Benford's Law. Datasets known to follow Benford's Law include financial data and population data, but also everything from scientific coefficients to baseball statistics and river lengths (Amiram, Bozanic, & Rouen, 2015; Diekmann, 2007; Hill, 1995) (Nigrini & Mittermaier, 1997).

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

The nature of expenditure data, which are based upon sums of numerous receipts that in turn include sums and multiplication of price times quantity, provide a theoretical basis for why we can expect Benford's Law to be the appropriate null hypothesis distribution for development expenditures. Appendix B.1 presents simulations showing that line-item totals, like the ones we analyze here, conform to Benford's Law. Moreover, across ecologically, economically, and demographically similar regions, we should expect similar patterns of digits when reporting is conducted honestly, even if Benford's Law did not hold.

3.2. *The statistical patterns of manipulated data*

Bureaucrats have an incentive to falsify expenditure data and embezzle both for personal gain and to satisfy kickback demands from superiors. Embezzlers weigh the costs and benefits of such behavior, including the probability of getting caught and the size of the penalty, in line with a rational decision to commit crime (Becker, 1968). In addition to prosecution, the costs of getting caught may include payoffs to

Table 1
Expected digit frequencies under Benford's Law.

Digit	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
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: This table shows the expected frequency of digits in each digit place according to Benford's Law (Nigrini & Mittermaier, 1997, p. 54).

auditors or others who detect their fraud, or career consequences imposed by their bosses. There may also be career consequences for refusing to participate in fraud perpetrated by one's superiors; this is especially common in systemically corrupt countries. Cheating may also be inhibited by personal or social values that provide disutility to dishonest behavior.

When a bureaucrat decides to fabricate data, we expect that they will manipulate the data to maximize payout and minimize the probability of detection. This can consist of a variety of behaviors. Bureaucrats falsifying reports are often subject to budget constraints for categories of expenditure but have flexibility over the value of each activity within that category; this was true in the World Bank project we analyze. Money can be skimmed either by adding line items that were never paid out (for example, ghost employees or trainings that never happened), or by padding the line items of genuine activities. Padding can take many forms, including over-invoicing schemes with contractors, in which case the outside party was aware, or by inflating the final expense in the report, which puts a premium upon keeping the reporting secret so that the contractors, beneficiaries, and other potential whistleblowers never know the official expenditure claimed for a project.[2] In line with a rational decision to commit fraud, we can expect that reporters increase data tampering in response to greater incentives to steal, and attempt to produce data that appear random to subvert detection. Furthermore, we expect that bureaucrats expend lower effort in subverting detection for data that are less likely to be monitored.

Bureaucrats who choose to produce false data face behavioral limitations on their ability to successfully do so. When experimental subjects are asked to produce random numbers, studies consistently show patterns of human digit preferences. In a study where students were asked to make up strings of 25 digits, their results followed neither the Benford distribution nor the uniform distribution (Boland & 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).

Evidence of specific digit preferences from Africa comes from an examination of African census data. A phenomenon known as "age heaping" occurs when people approximate their age: 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 specific digit preferences. 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, & Snavley, 1973; UN Economic and Social Council Economic Commission for Africa, 1986). These same digit patterns occur in our data; both 0 and 5 are so heavily overrepresented that we analyze their usage separately and analyze only digits 1–4 and 6–9 in most of our analyses. Nevertheless, we can rule out the idea that the patterns present in our data are the result of legitimate digit preferences for underlying *prices*. Appendix B.1 presents a simulation where underlying prices are contaminated with digit preferences, and yet line-item totals, like the ones we analyze here, still conform to Benford's Law.

4. Data

4.1. World Bank expenditure and participant data

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 11 arid districts and 17 semi-arid districts that were added after 2003. This community driven development project spent $224 million USD targeting the most impoverished people in the heavily drought-prone regions of Kenya. It funded small infrastructure (such as schools, dispensaries, and water systems), income-generating activities (such as goat restocking), drought and natural resource initiatives, and training exercises for villagers. In Appendix A we also present extensive qualitative data about the mechanisms by which corruption operated in the Arid Lands project.

Our digit analysis is confined to the 11 arid districts, as these districts were the most homogeneous across ecological, economic, and demographic measures. The expenditure and participant data used in these analyses were extracted from quarterly project reports produced by each district. These reports break out the expenditures and numbers of male and female participants associated with most activities undertaken by the project in a given district and year. Each line-item expenditure represents the total expenditures for that project, for example, a classroom, a goat restocking project, or a well rehabilitation. The ability to perform analysis on both participant and expenditure data is valuable, as it allows us to compare patterns that arise when the same individuals manipulate very different numbers.

These districts were all subject to the same project rules and the same level of monitoring. They also share many similar characteristics: their economies depend primarily upon livestock, they are among the poorest and most drought-prone in Kenya; they are remote from centers of power, sparsely supplied with infrastructure (roads, schools, health services, access to clean water, and electricity); and their populations are poorly educated. These similarities are important because they allow us to assume that there were no legitimate reasons to expect differences in digit patterns across districts.

Table 2 presents a full statistical description of the data used in this project. The expenditure values (in Kenyan shillings) range from a minimum of 1,508 to a maximum of 9,000,000. The mean expenditure is about 268,000 shillings and the median is about 124,000 shillings. We have 4,339 expenditure observations; note, however, that the sample size of the statistical tests we do is much higher than 4,339, because each number contains multiple digits. We have 5,499 observations of beneficiaries, where one observation contains data on a project, with male and female beneficiaries usually listed separately.

Our data contain 11 geographic districts and 4 project periods (years). We see some variation in sample size per district, with a mean of 394 expenditure observations per district and a standard deviation of 91, a minimum of 293 expenditure observations (Marsabit district) and a maximum of 578 expenditure observations (Wajir district). Among years, we observe data from 2003 to 2006 (a single "year" for reporting purposes), 2007, 2008 and 2009. We see roughly equal sample sizes between years, with between 944 and 1249 expenditure observations per year.

4.2. Forensic audit data

In 2009, following an external complaint, the World Bank's Integrity Vice Presidency (INT) began a forensic audit of the Arid Lands project that lasted 2 years and culminated in a public report (World Bank

[2] There was a premium placed upon keeping reporting data private in this project, even from other high-level project officers working in the same district office. One of the authors spent 2 years negotiating with the World Bank for access to these reports and was granted access only after intervention from the U.S. representative on the Board of the Bank on the grounds that the original project document promised that these data would be made public (World Bank, 2003). Even so, only about 2/3 of the reports were ever released.

Table 2
Descriptive statistics of World Bank Arid Lands expenditure and participant data.

Numeric Variables	Min, Max	Interquartile Range	Mean	Median	Sample Size
Expenditure	[1508, 9000000]	[56000, 278000]	267,925	123,984	4399
Male Beneficiaries	[0, 989]	[16, 90]	93.84	40	5232
Female Beneficiaries	[0, 999]	[11, 80]	85.91	33	5207
Categorical Variables	**Unique Values**	**Min & Max # Expenditure Obs per Category**	**Mean # Expenditure Obs per Category**	**Median # Expenditure Obs per Category**	
Districts	11	[293, 578]	394.5	359	
Years	4	[994, 1249]	1084.8	1073	
Sectors	5	[127, 2667]	867.8	464	

Notes: This table presents summary statistics of the categorical and numeric data used in our analysis. The top panel describes the distribution of our numeric variables subject to digit analysis, and the bottom panel describes the categorical variables by which we break out our analysis.

Integrity Vice Presidency, 2011).[3] Auditors sampled 2 years' worth of receipts (2007–2009) for 7 districts, 5 of which (Wajir, Isiolo, Samburu, Garissa, and Tana River) were arid districts examined in this analysis. They examined the underlying supporting documents for 28,000 transactions, which are equivalent to our line-item project data. The auditors worked from actual project receipts and supporting documents, such as cashbooks, bank statements, and vehicle logs. They also travelled to the districts to conduct interviews with suppliers to verify the legitimacy of suspicious transactions. The outcome measure we use for this comparison to our own results is the percentage of suspected fraudulent and questionable transactions by district.

5. Digit tests and results

We provide a set of 10 non-overlapping tests that capture different ways in which data can be manipulated. Table 3 lists these tests. Two of our tests are new, and two more build upon our new tests to show specific examples of data manipulation. The remaining tests are variations on existing tests in the literature. We collect the findings of all 10 tests together and compare against those of the World Bank forensic audit. To account for multiple tests, we use a Bonferroni correction: we divide our desired significance level (0.05) by the number of tests (10) and set a significant level of $p = 0.005$, used throughout our analyses. The summary of our tests' statistical significance is presented below; full details of the p-value and sample size for each test are provided in Section 5.6.

To facilitate the use of this method by other researchers, we have made this code available as an R package called digitanalysis, which can be accessed online at http://github.com/jlederluis/digitanalysis.

Table 3
List of non-overlapping tests performed.

Test Description	Section
New Tests	
1) All Digit Places Beyond the First—Expenditure Data	5.1
2) All Digit Places Beyond the First—Participant Data	5.1
3) Padding Valuable Digit Places	5.2
Applications of New Tests	
4) Unpacking Rounded Numbers	5.3
5) Election Year Effects	5.4
Adaptations of Tests from the Literature	
6) Rounding	5.5.1
7) Repeated Numbers	5.5.2
8) Differences Across Sectors	5.5.3
9) First Digits	5.5.4
10) Digit Pairs	5.5.5

5.1. All digit places beyond the first

A simple, yet powerful, test of data manipulation is conformance of the observed digits to Benford's Law. Such tests are frequently performed in a single digit place, using the first, second, or last digit place (Diekmann, 2007; Beber & Scacco, 2012). In this new test we examine multiple digit places simultaneously. Compared with single-digit-place tests, a simultaneous analysis of multiple digit places increases sample size for statistical testing and therefore vastly increases statistical power. The increase in sample size afforded by simultaneous-digit-place analysis is especially helpful when analysis can benefit from data disaggregation, which can result in low sample size. Furthermore, testing individual digit places results in multiple hypothesis testing issues, which a simultaneous test of all digit places avoids. Additionally, we omit the first digit when conducting this analysis, because individuals tampering with data may not have complete control over the leading digit or may avoid changing it to subvert detection. This has the potential of a more powerful fraud detector because the noise of the first digit, which may have been left clean strategically, is eliminated. The first digit test alone is presented separately below.

We use a two-way chi square test to compare the contingency table of all digit places beyond the first against the Benford distribution. We omit 0 and 5 from this analysis, which may be subject to rounding for legitimate reasons, and which we handle separately below in a test for excess rounding. We correct for the removal of 0 s and 5 s by renormalizing the expected frequencies of the remaining digits, so the total expected probability sums to 100 %. For each digit place (2nd digit, 3rd digit, etc.), the frequency of each digit (1, 2, 3, 4, 6, 7, 8, and 9) is compared with the expected frequencies given in Table 1.

Figures 2AB present the data of all digit places beyond the first for expenditure (Panel A) and participant data (Panel B). Data from all digit places are projected onto one axis for visualization. Among the expenditure data for all districts in Figure 2, Panel A, we see a strong preference for digits 2 and 8, underreporting of 1 and 9, and overall nonconformance to the expected Benford distribution ($p = 3.9 \times 10^{-15}$). Strikingly, these same digit patterns appear even more strongly in the participant data (Panel B), and the result for all district data combined is again highly significant ($p = 5.7 \times 10^{-51}$). This pattern is also consistent

[3] The World Bank referred the Arid Lands case to the Kenyan Anti-Corruption Commission after completing a joint review together with the Kenya National Audit Office, which confirmed the findings and resulted in the Kenyan government's agreement to repay the World Bank \$3.8 million USD for disallowed charges (World Bank Integrity Vice Presidency and Internal Audit Department, Treasury, Government of Kenya, 2011). It is noteworthy that the Kenyan Anti-Corruption Commission did not follow up and no one from the senior management in headquarters was prosecuted or fired. Such impunity is common in systemically corrupt countries and speaks to the need for donors themselves to be more vigilant. The World Bank did refuse to renew the project in 2010, even though it already had a Board date set for a 5-year renewal.

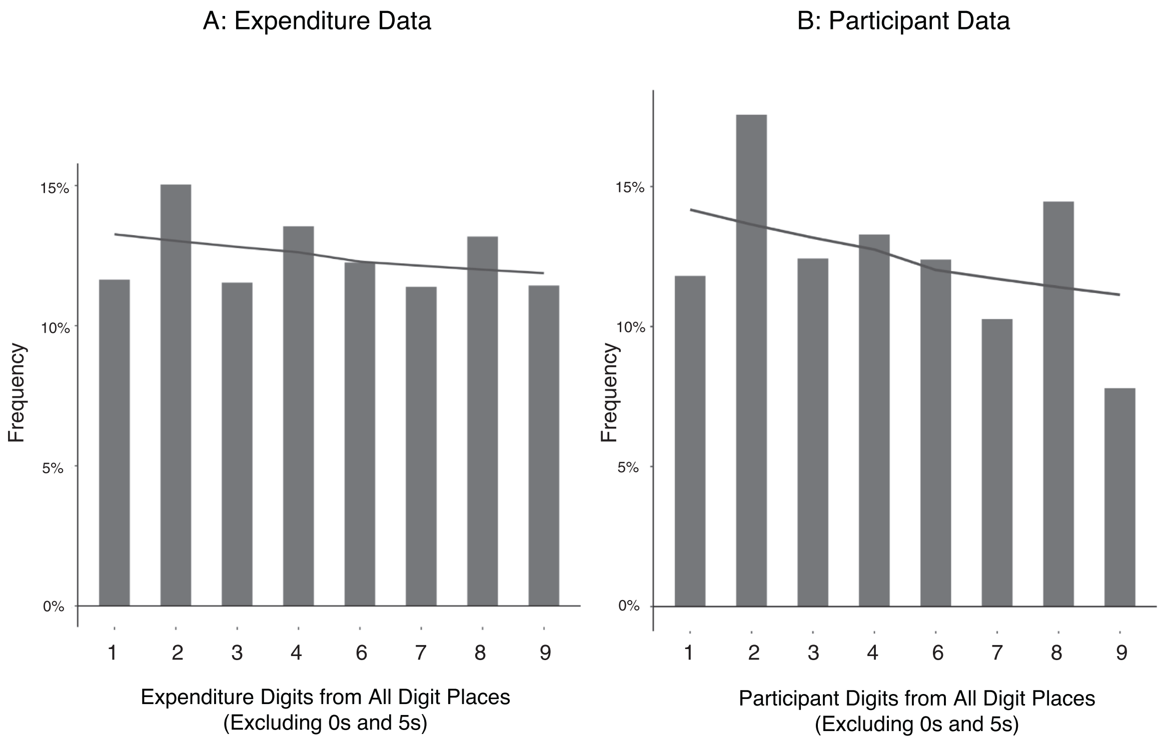

Figure 2. All digit places beyond the first vs Benford's Law for expenditure and participant data.
Notes: This figure presents all digits from beyond the first place from expenditure data (Panel A) and participant data (Panel B) for all districts combined. The expected Benford's Law distribution is the solid line. Both tests are statistically significant, with $p = 3.9 \times 10^{-15}$; $n = 9371$ for the expenditure data (Panel A) and $p = 5.7 \times 10^{-51}$; $n = 7385$ for the participant data (Panel B). Notably, both datasets show preferences for even numbers, particularly 2 and 8. The digits 0 and 5 are omitted due to heavy overuse that may be legitimate rounding.

with the humanly generated African census pattern (Nagi, Stockwell, & Snavley, 1973); 2 s and 8 s are high, and 1 s and 9 s are low. In 8 of our 11 districts, we reject the null hypothesis that all digit places conform to Benford's Law for both the expenditure data and the participant data at the $p < 0.005$ level.

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.

Appendix B.2 presents simulations comparing the power of all digit places testing to single-digit testing and shows that it has a much higher rate of true-positive detection

5.2. Padding test for strategic manipulation in valuable digit places

Basic tests of conformance to Benford's Law, including our own all-digit places test, are not sensitive to the magnitude of the values of manipulated data. This is a major limitation of traditional Benford's Law testing. Evidence that data are being fabricated consistently in the direction of increasing payment to the embezzlers is important evidence of intent, which is a critical component to the distinction between fraudulent manipulation and accidental errors. While there may be a strong correlation between incomplete paperwork and actual embezzlement, it is not necessarily the case that sloppy bookkeepers are misappropriating funds. This may be even more relevant in the developing world where staff are likely to be less well educated. For this reason, evidence that points to consistently profitable deviations from expected digit distributions, or evidence of strategic efforts to avoid detection, bring us a step closer to showing intent to defraud.

As discussed in Section 3.2, bureaucrats falsifying data can be expected to inflate values to receive greater illicit reimbursement. 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.

Benford's Law governs the distribution of digits by the number of positions from the left (1st digit, 2nd digit). However, the value of a digit depends on the digit's position from the right (e.g., 1 s, 10 s, 100 s place), and this value determines the incentive to manipulate a digit.

To overcome this limitation, we compute the expected mean under Benford's Law by digit place from the right (10 s, 100 s), using the length of the numbers in our dataset to match left-aligned digit places and right-aligned digit places. We consider 5-, 6-, and 7-digit numbers, to ensure sufficient sample size in each digit place, and drop 0 s and 5 s, which are handled separately when we check for rounding. In each digit place from the right (1 s, 10 s, etc.), we compute the Benford expected mean as follows: for 5-digit numbers, the Benford mean in the 10,000 s place is the mean of the 1st digit; for 6-digit numbers, the Benford mean in the 10,000 s place is the mean of the 2nd digit; etc. We compare the observed mean of our data to the expected mean under Benford's Law. This is a difference of means statistic, for which a positive value indicates a mean greater than the expected mean under Benford's Law.

To determine significance of each of our statistics, we perform a Monte Carlo simulation. We generate 100,000 observations of means drawn from the Benford distribution for the appropriate digit place. We remove 0 s and 5 s and compute the means by digit place from the right as well as the Benford expected mean, identically to the way we process the real data. For each of the 100,000 observations, we produce a difference of means statistic. We then compare our observed difference of means statistic from the data to these simulations. The *p*-values reported are the empirical cumulative distribution function (CDF) of our difference of means among the simulated statistics. That is, if our statistic exceeds 90 % of the simulated values, its p-value is 0.10. For a simulation with K samples, there is a minimum p-value of 1/K.

Figure 3 shows the padding tests for both World Bank and simulated data against the Benford expected distribution. The 0 line indicates the

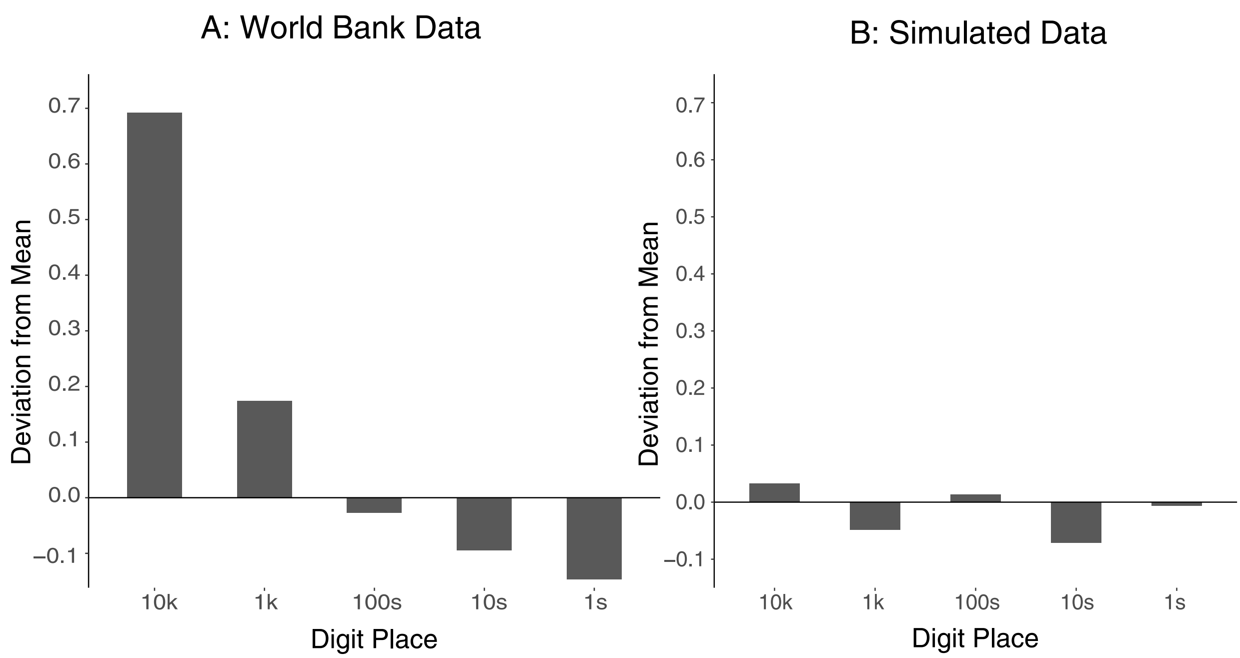

Figure 3. Padding test of valuable digit places vs. simulation.
Notes: We compare the mean in each digit place from the right to the Benford expected mean. Zero reflects conformance to the Benford expected mean, and positive values indicate the mean digit is higher than Benford's Law predicts. The observed pattern in the World Bank Data (Panel A) is consistent with an intentional strategy of placing high digits in high digit value places and then underusing them in low digit value places to even out the digit distribution. Compared to a sample of 100,000 Benford-conforming simulations, we observe the following statistics for the World Bank Data: 10,000 s place ($p = 1.0 \times 10^{-5}$), 1,000 s ($p = 2.3 \times 10^{-4}$), 100 s ($p = 0.33$), 10 s ($p = 0.10$), 1 s ($p = 0.061$). Panel B shows the simulated Benford-conforming data with 10,000 observations. No such pattern emerges.

Benford mean; anything above the line represents an overuse of high digits, and anything below the line represents an underuse. The World Bank project data (Panel A) in the 10,000 s place exceed 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. Ten thousand Kenyan shillings was worth approximately $150 USD in 2007.

Perhaps the most interesting finding in Figure 3A, which points to intention to conceal, is the decline in the use of high digits as one goes from the 10,000 s to the 1,000 s, 100 s, 10 s, and 1 s places. 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 presence of too many high digits in any given table in the report. In contrast, Figure 3B, which uses simulated data that conform to Benford's Law, show no such pattern, and the deviation from Benford's Law is randomly distributed around 0.

In sections 5.3 and 5.4 we provide examples of how our two new tests can be applied to reveal the effects of behavioral limitations (all digit places but the first) and political incentives (padding valuable digit places).

5.3. 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 at 100 Kenyan Shillings (about $1.50 USD) per person, per day in 2007. It is reasonable to assume that the authors of the annual district reports expected that participant data would be less scrutinized than 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. Second, training exercises in remote villages are difficult to verify because their final product is human capital, which leaves no physical evidence. 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, undertook low-effort fabrication, and reported a round total number of participants to meet that target. The total number of participants was then split into males and females, as was required for reporting, and consequently hid the presence of round numbers. Therefore, we expect greater indicators of data fabrication when the total number of male and female participants sums to a round number.

To test this, we analyze the distribution of all but first digits of reported numbers of total participants (males and females) when their sum ends in a 0 versus a non-0 digit. We perform the multiple-digit-places-test on these two samples, as an application of our new method, using all digits beyond the first. 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 2 s and 8 s and low incidence of 1 s and 9 s when the gender specific data come from a pooled number that ends in 0 (Figure 4A, left). This pattern is consistent with humanly generated data and not with naturally occurring data. There is still evidence of human generation in the data when the gender total is not round, Figure 4A right ($p = 1.9 \times 10^{-6}$), but the statistical significance is even higher in the rounded data, Figure 4A left ($p = 2.6 \times 10^{-64}$ in the sample of all districts). For 8 out of 11 districts, we reject the null hypothesis that the total of male and female participant data is Benford conforming ($p < 0.005$).

The validity of this test hinges on the fact that, under Benford's Law, data from two Benford distributions where the sums happen to end in a round number still follow the Benford distribution. This is not a trivial idea; it is possible that, by conditioning on the sum of two numbers drawn from Benford distributions, the digits of the data that produce that sum have some legitimate reason to come from a different distribution.

To validate this, we simulate independent Benford conforming "male" and "female" participant values between 2 and 4 digits and sum

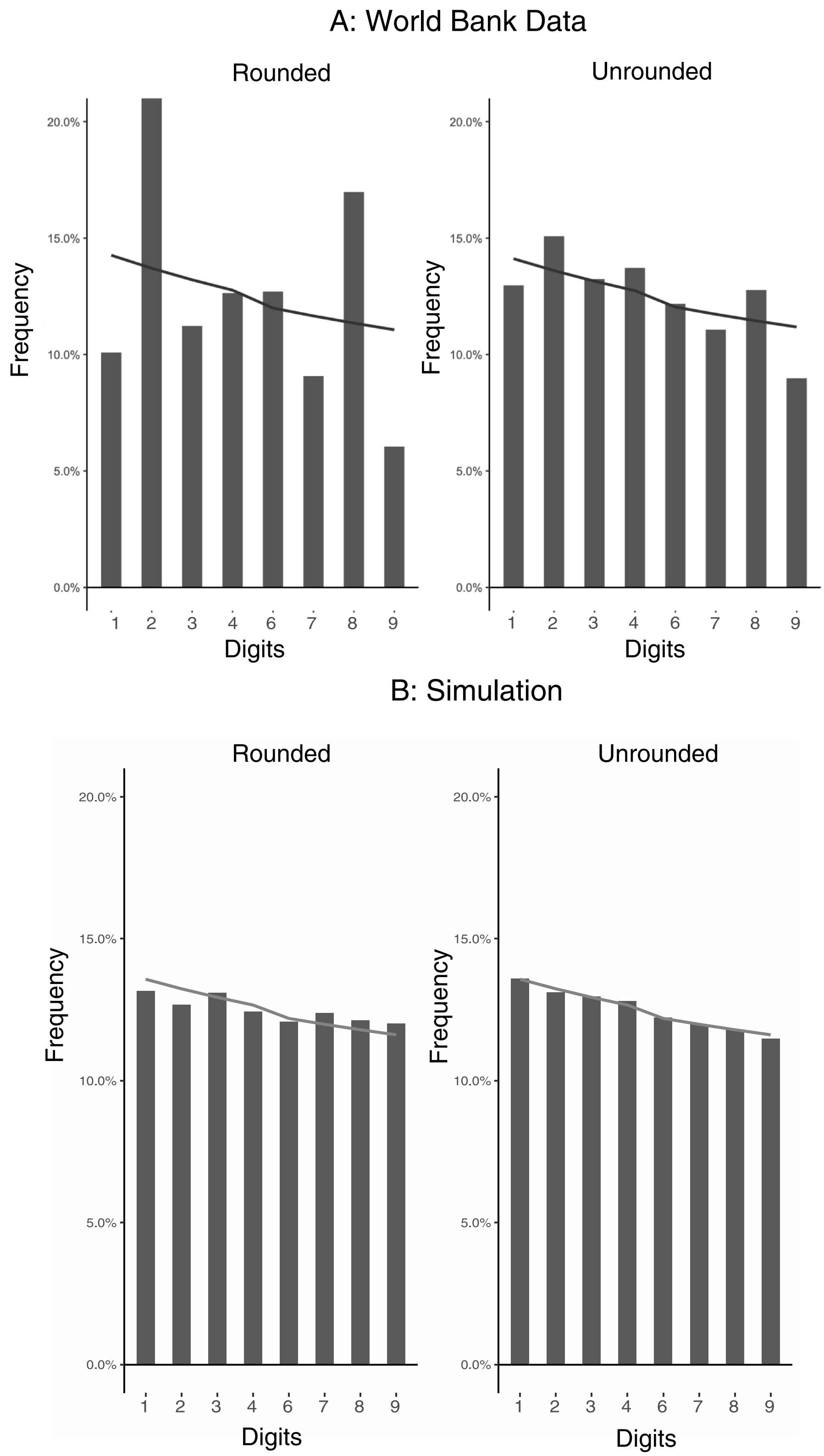

Figure 4. Unpacking rounded and unrounded digits in participant data.

Notes: PANEL A: World Bank Data. This presents a test of all digit places beyond the first digit among participant data (male and female pooled), when the total of male and female participants sums to a rounded number or an unrounded number. In the World Bank Data (Panel A), data that sum to a round number show higher preferences for even numbers, although both samples fail tests of conformance to Benford's Law: rounded data, $p = 2.6 \times 10^{-64}$; $n = 2975$, unrounded data, $p = 1.9 \times 10^{-6}$; $n = 4410$. PANEL B: We compare this to a simulation of $n = 50{,}000$ observations, where male and female numbers are generated independently in conformance with Benford's Law and then summed, and we analyze sums that happen to be rounded versus those that do not. The simulation is not statistically significantly different from Benford's Law, $p > 0.01$, and there are similar patterns between rounded and unrounded data.

them. We then condition on whether that sum is rounded or not. Panel B of Figure 4 shows the result of this simulation. We find no divergence from Benford's Law evident in simulated data; both the left and right panels (totals ending in 0 or not) show conformance to Benford's law. This is evidence that the patterns found in the World Bank Data (Panel A) are the result of human manipulation.

This test highlights the power of our all-digit places analysis. Pooling digit places increases sample size, allowing analyses that can partition data along different categories to capture behavioral patterns that may not arise when examining data in aggregate.

5.4. Election year effects

Interview data frequently cited the connection between syphoned project funds and the controversial presidential political campaign of 2007. The association between corruption and political campaigns has also been noted in other studies (Claessens, Feijen, & Laeven, 2008). The next test partitions our data by project year to examine whether the evidence is consistent with higher rates of embezzlement in the presidential election year 2007. We look for padding of high-digit numbers by project year by using our new padding test, with expenditure data disaggregated by year. We compare 2007 to the Benford-conforming baseline and repeat our Monte Carlo statistic by year. Relatedly, in forensic accounting, auditors may examine the time-dimensionality of irregular expenditures, and recent work has shown the value of such analyses in detecting corporate accounts misreporting (Cheng, Palmon, Yang, & Yin, 2022).

As we see in Figure 5, in 2007 (the only presidential election year) there was a statistically significant overuse of high digits in valuable digit places ($p = 0.001$). This is consistent with a greater incentive to embezzle to support political campaigns during the highly controversial presidential election year that led to extreme violence (Gibson & Long, 2009).

5.5. Other tests from the digit analysis literature

We present the results of 6 other tests that also exhibit the behavioral limitations and economic incentives expected from fabricated data. These tests are standard in existing digit analysis literature and include tests for first-digit conformance to Benford's Law, rounding of numbers, repeated data, increased rounding in lesser monitored expenditures, the underuse of "digit pairs" (e.g., 22), and last digits. These tests all corroborate that the World Bank data are highly manipulated and allow us to examine different signals of this behavior graphically and statistically.

5.5.1. Rounding

It is common for auditors to look for both high levels of rounded and

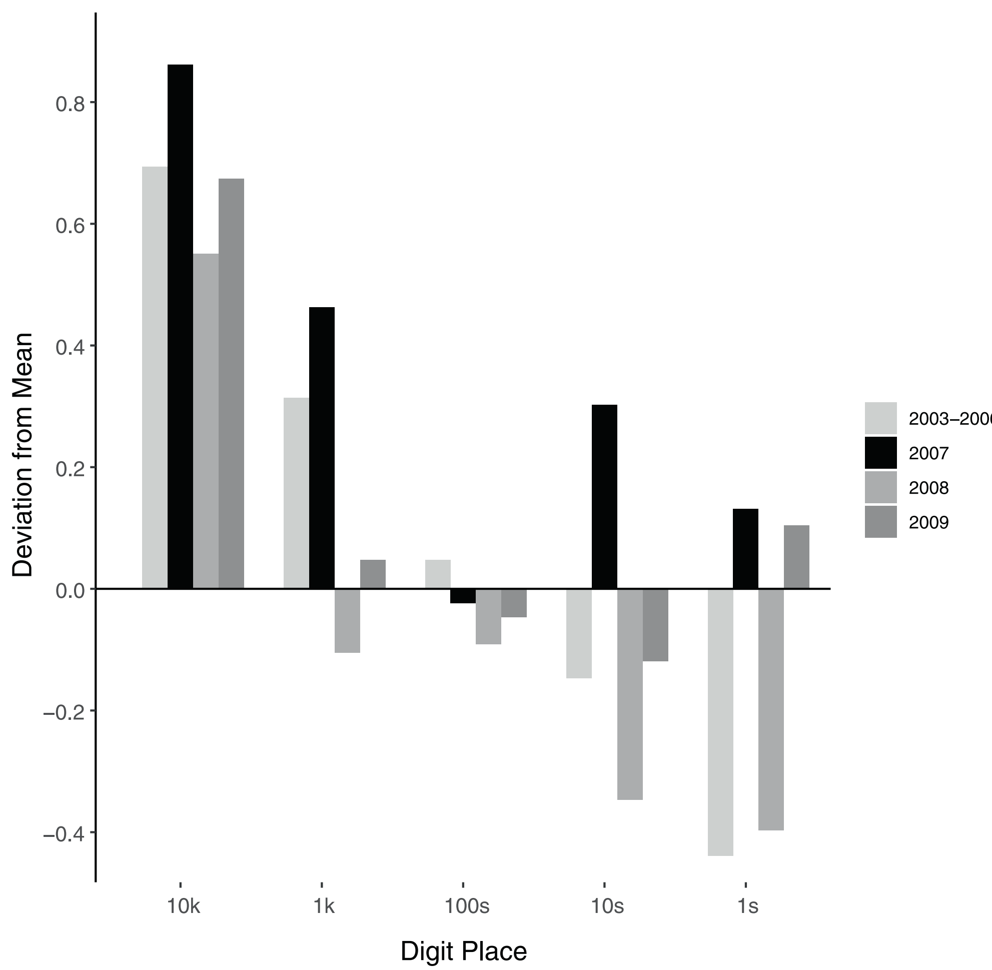

Figure 5. Election year effects in expenditure data.
Notes: This figure performs the padding test by year. 2007 was a Presidential election year and has a statistically significant overuse of high digits in valuable digit places, even more than other years, (ten thousand place, $p = 0.0001$; one thousand place, $p = 0.0001$).

repeated data, and these are often viewed as potential evidence of human tampering (Nigrini & Mittermaier, 1997) (Chin Hsien & Lin, 2011). In the absence of theoretically acceptable levels of rounding and repeating, we compare districts to each other, as there is no known reason to expect differences among such ecologically, economically, and demographically similar districts.

The Kenyan shilling exchange rate was 66 Kenyan shillings to $1 USD in 2007. Its value was low enough that many receipt data would legitimately show high levels of 0 s and 5 s in the terminal digit place. However, one must bear in mind that 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 entire transaction, and it is the full transaction totals that we are examining.

We count the number of rounded digits, tallying the number of trailing 0 s (0, 00, 000, etc.), or digits in terminal strings of 5, 50, or 500, as a fraction of the number of digits in each line item. For example: the number 30,000 has 4 rounded digits out of 5 (80 %); the number 12,350 has 2 rounded digits out of 5 (40 %); and the number 11,371 has 0 rounded digits. Rather than counting 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). Figure 6 shows the average percentage of rounded digits by district.

While we don't know the level of rounding that would occur naturally in an honest dataset, there is good reason to expect that the same type of retailers, servicing the same type of contracts in economically, ecologically, and demographically similar districts, practiced the same rates of rounding. In the absence of an expected level of rounding, we compare districts to each other. For each district, we conduct a Welch's unequal variances *t*-test to compare the mean percentage rounding to all other districts. For example, the statistical test for Baringo compares the level of rounding in Baringo to the level of rounding in the 10 other districts combined. We conduct a one-tailed test to check for excessive rounding and define statistical significance at $p < 0.005$. Four of the districts fail this test.

5.5.2. *Repeated numbers*

Exactly repeated numbers are also a red flag for auditors (Nigrini & Mittermaier, 1997; Debreceny & Gray, 2010) (Knepper, Lindblad, & Seifu, 2016). Our hypothesis is that embezzlers expended less effort in data fabrication when there was less reason to expect scrutiny. Repeated values are consistent with low-effort data fabrication. One such example is remote training exercises, which are particularly hard to verify.

A specific example from the Tana District Report of 2003–6 illustrates the problem of repeated data (Republic of Kenya, 2006). On page 49, we find 8 training exercises listed that took place in different villages for 3 weeks, each from March 5–27. The district had neither enough vehicles, nor enough training staff to run 8 simultaneous trainings. Among the 8 expenditures listed, we find the identical cost (245,392 Kenyan Shillings) listed for 3 different trainings, and another number (249,447) exactly repeated twice. Trainings are the summed costs of the per diems for 4–5 trainers and 1 driver (at different rates), the cost of fuel to the destination, stationary for the seminar, and 100 Kenyan Shillings 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; indeed, the cost of food alone for 172 trainees should have exceeded all the amounts listed.

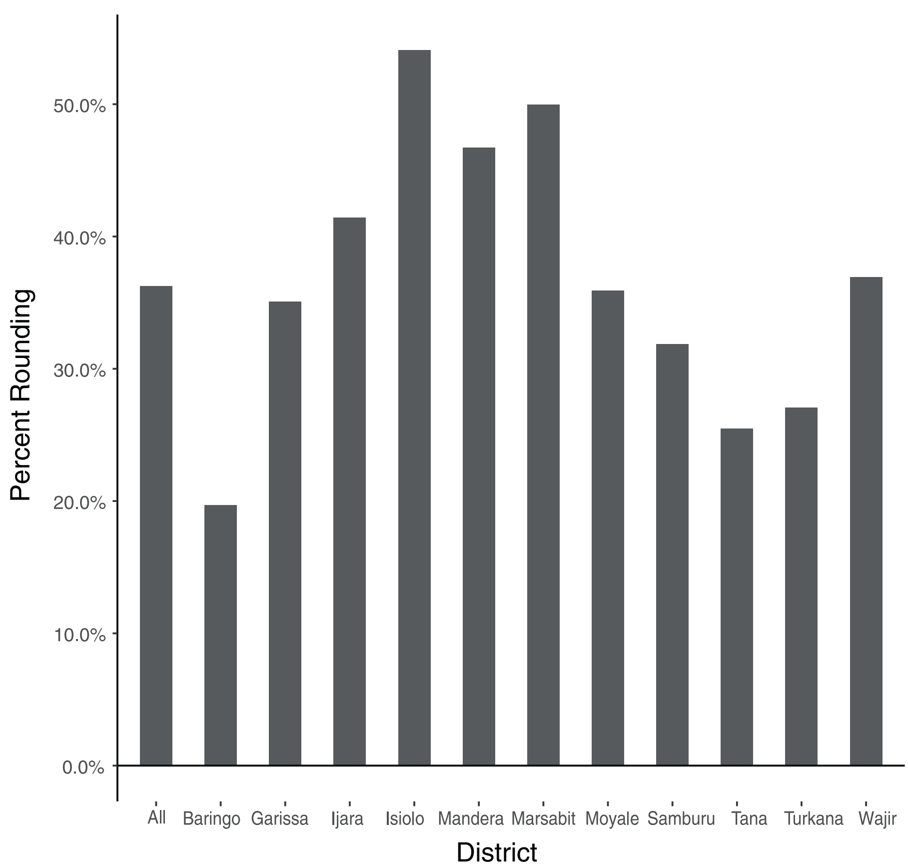

Figure 6. Percentage of rounded digits in expenditure data by district.
Notes: This figure shows the percentage of digit places rounded in expenditure data by district. For each district, we compare the level of rounding to the level in all other districts and conduct a one-tailed *t*-test for excessive rounding. Ijara, Isiolo, Mandera, and Marsabit are statistically significant in their overuse of rounding as compared to other districts ($p < 0.005$).

In our calculations, repeating numbers refer to the use of identical expenditure amounts for different activities. We define an exact repeat to be an expenditure matching year, district, sector, and expenditure value. There is no correction for rounding in the repeating data, as we wish to maintain the independence of our tests for rounding and repeating.

Figure 7 shows the results for the percentage of line items that repeat exactly. As with rounding, the empirically truthful level of repeating is unknown but there is no reason for patterns across districts to differ. We compare each district's average amount of repeated numbers to all other districts, using a Welch's unequal variance *t*-test, and conduct a one-tailed test for excessive repeating as compared to all other districts. Three districts (Baringo, Isiolo, and Mandera) fail this test. We also see suspiciously wide variation across districts: Baringo approaches 50 percent, while Turkana repeats about 5 percent. Notably, our rounding and repeating tests flag different districts, indicating that they pick up different signals.

5.5.3. *Differences across sectors*

Economic theory (Becker, 1968) and empirical work (Olken B. A., 2007) indicate that individuals are more likely to cheat when there is a lower risk of detection. The training and transport sectors of this project (travel, fuel, and vehicle maintenance) provided greater opportunities for individuals to pad expenditures relative to the civil works and goods and equipment sectors. While the latter left physical evidence of spending (such as a classroom), the former did not. For example, tracking down nomads who were reported as present for a training exercise in a remote village 2 years prior to an audit is all but impossible. Similarly, project fuel could have been diverted to private vehicles while leaving no trace. Therefore, we predict that individuals fabricating data for these sectors may have done so with less effort expended on deception. To detect this, we look for evidence of a greater incidence of repeated numbers among training, travel, and vehicle expenditures. We plot the percentage of repeated line items that match year, district, and amount, for each of the districts by sector in Figure 8.

For each district, we conduct a Welch's unequal variance *t*-test of the number of repeats in the training and transport sector versus the civil works and goods and equipment sectors from the same district combined. Seven of 11 districts and the all-district test have statistically higher repeats in that sector. In Turkana, Garissa, and Tana River Districts, other sectors have higher percentages of repeats, providing evidence that there is no structural reason for this phenomenon. While we don't know what the empirically honest level of repeating should be, there is no known legitimate reason for there to be more repeated line items in some districts than others. This test differs from the simple test of repeats because the sector test compares differences in repeating within a district, with the assumption that repeating should be constant across sectors.

5.5.4. *First digits*

We perform a test of the first digit place, which is common in the digit analysis literature (Durtschi, Hillison, & Pacini, 2004) (Nigrini, 2012). In the first digit place, we expect digits to follow (Hill, 1995):

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

Figure 9A plots this distribution as a solid line and shows the conformance of the first digits to Benford's Law. Data from the full sample of districts 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 necessarily mean that all the first-digit data are unmanipulated. First, people may resist tampering with the first digit to avoid detection. Second, pooled data may cancel

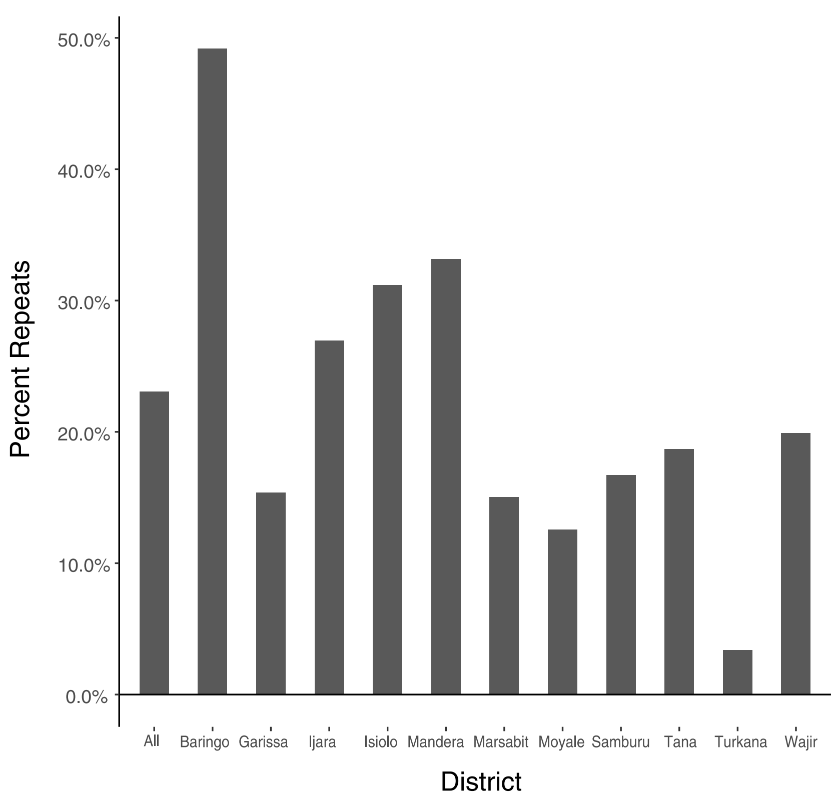

Figure 7. Percentage of repeated numbers in expenditure data by district.
Notes: This figure shows the percent of exactly repeated expenditure entries by district for a given annual report. For each district, we compare the level of repeating to the level in all other districts and conduct a one-tailed *t*-test for excessive repeats. Baringo, Isiolo, and Mandera are statistically significant in their overuse of repeating compared to other districts ($p < 0.005$).

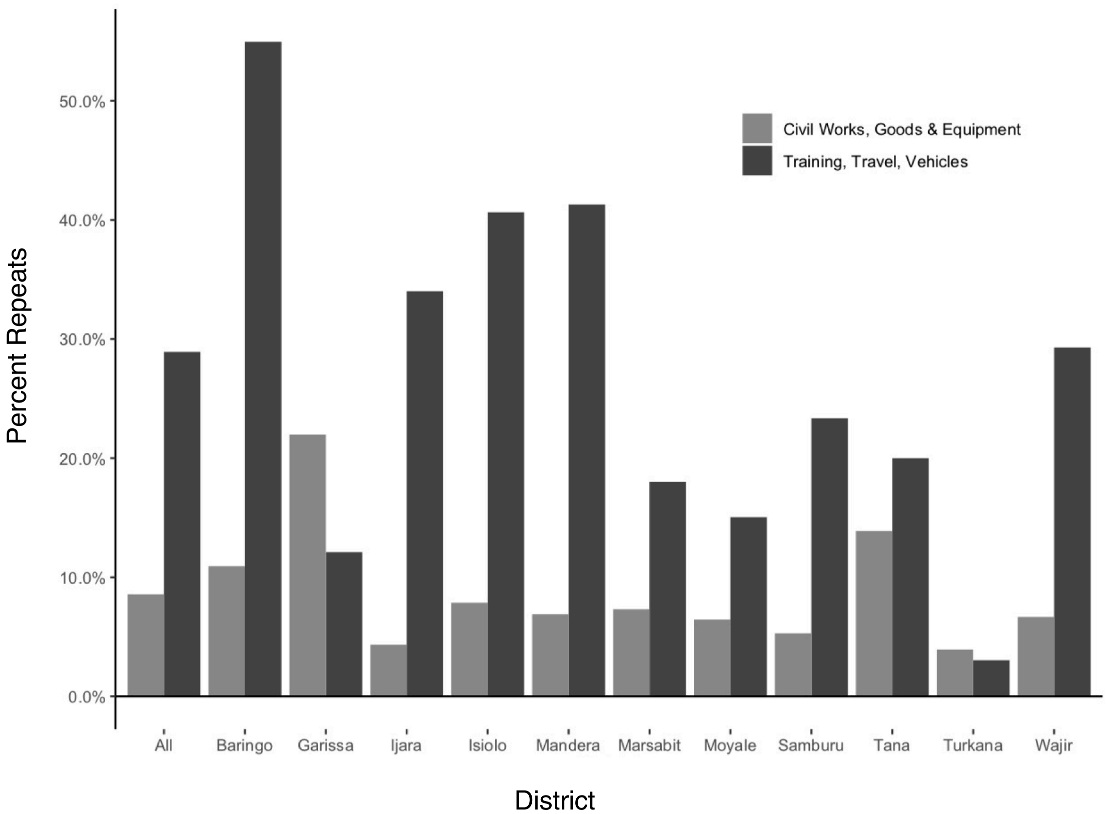

Figure 8. Sector effects in expenditures.
Notes: This figure plots the percentage of line-item expenditures repeated exactly, matching on district, year, and sector. We test whether harder-to-verify expenditures from training exercises, travel, and vehicles are more likely to be repeated than expenditures in civil works projects and purchases of goods and equipment. The districts of Baringo, Ijara, Isiolo, Mandera, Marsabit, Samburu, Wajir, and all districts combined show statistically significantly higher repeats ($p < 0.005$).

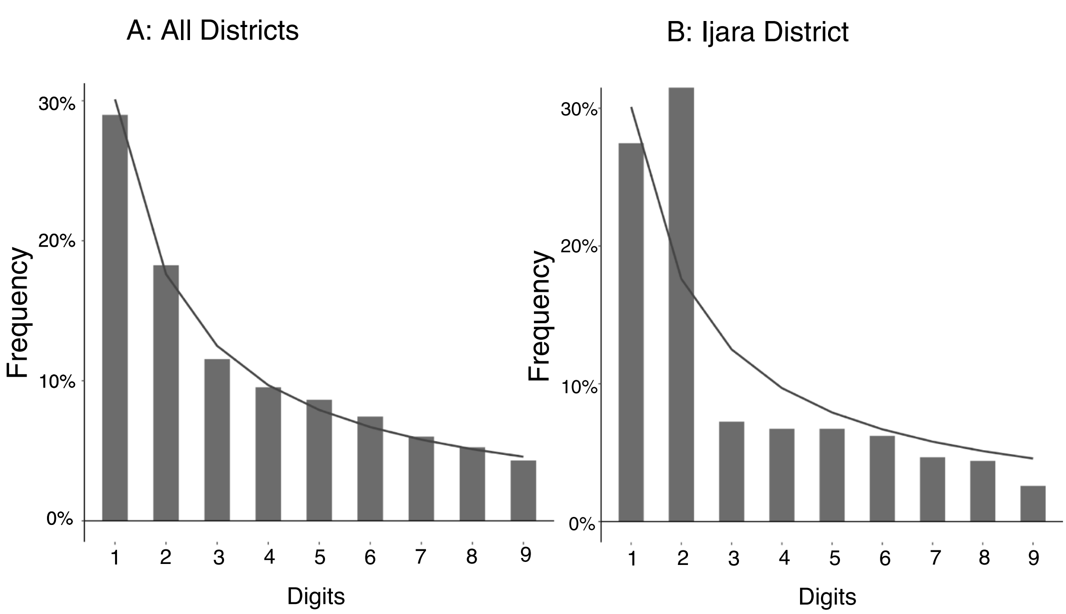

Figure 9. First-digit expenditure data against Benford's Law.
Notes: PANEL A: This figure presents the first-digit test as compared to Benford's Law for All districts combined ($p = 0.089$; $n = 4339$). PANEL B: Ijara District only ($p = 2.3 \times 10^{-13}$; $n = 386$). The line represents the expected distribution under Benford's Law. While the aggregate data from all districts conform to Benford's Law, 7 or our 11 districts did not.

out different individual signatures of manipulation and replicate Benford's Law (Diekmann, 2007). Appendix B.1.2 reports simulations that exhibit this later phenomenon; overall data conform to Benford's law, while data disaggregated by reporter may not. This becomes evident when we look at the data from individual districts where the reports were constructed. Figure 9B 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.

5.5.5. Digit pairs

Underuse of digit pairs, e.g., 11, 22…99 in adjacent digit places, is a common feature of humanly produced data (Boland & Hutchinson, 2000; Chapanis, 1995). Other applications of digit analysis examine the last 2 digits (Nigrini M. J., 2012), or explicitly test for digit pairs (Beber & Scacco, 2012; Adiguzel, Cansunar, & Corekcioglu, 2020) (Barney & Schulzke, 2016).

Among the participant data, we expect a uniform distribution of terminal pairs, 9 of 99 pairs. We omit the pair 00 in case it is affected by rounding. We 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. These data most typically record the number of women and men (listed separately) who showed up in response to an open invitation to appear for a training exercise in their village. 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.

A digit pair analysis of participant data is shown in Figure 10. Five of the 11 districts significantly underuse final-digit pairs in the participant data at $p < 0.005$ significance, as does the combined sample of all districts ($p = 2.5 \times 10^{-10}$).

Due to the low value of the Kenyan shilling, rounding in the last digit places may be legitimate in expenditure data. Therefore, an equivalent analysis of expenditure data is not appropriate. For this reason, we confine our analysis to the beneficiary data, where there is no legitimate reason for rounding in the ones place, as participant data are reported as exact counts of people who show up.

5.5.6. Last digits

Literatures on both forensic auditing and election fraud emphasize analysis of terminal digits, which should be uniformly distributed if they represent the fourth digit place or beyond (Nigrini & Mittermaier, 1997; Beber & Scacco, 2012).

Results on the terminal digit show exceptional statistical significance for both expenditure and participant data; we present these in Figure 11. Both expenditure data and participant data are statistically significant against Benford's Law when combining all districts, with $p < 0.005$. We exclude this test from our aggregate analysis below because last digits are subsumed in our test of all digit places. Appendix B compares the all-digit places test to single-digit tests including this last-digit test and concludes that simultaneous testing has better statistical power.

5.6. Summary of tests

Table 4 compiles the results of all 10 non-overlapping tests for each district. To address type 1 error due to the number of tests we conduct, we perform a Bonferroni correction and divide our desired significance level (0.05) by the number of tests (10). This sets a significance level of 0.005. These 10 tests avoid overlap and pinpoint different aspects of data manipulation. The bottom row shows the number of tests that show statistically significant deviation by district, which averages 5.7 out of 10.

6. Validity

We establish the validity of our statistical testing method in two

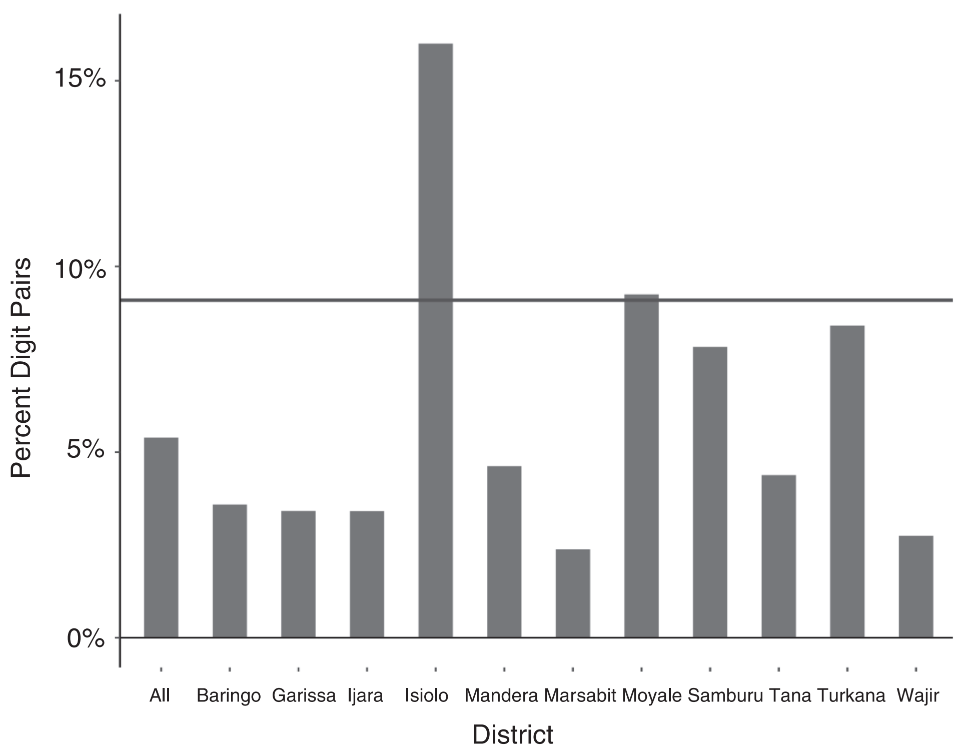

Figure 10. Digits pairs in the last two digits for participant data by district.
Notes: We test for underuse of digit pairs such as 11, 22, and 33 in the last two digits. Baringo, Garissa, Ijara, Marsabit, Wajir, and all districts underuse digit pairs under a binomial test with $p < 0.005$. The line represents the expected distribution of digit pairs under the uniform distribution, where 9 out of 99 substrings should be repeated values, (omitting the rounded substring 00 which is tested separately in Figure 6).

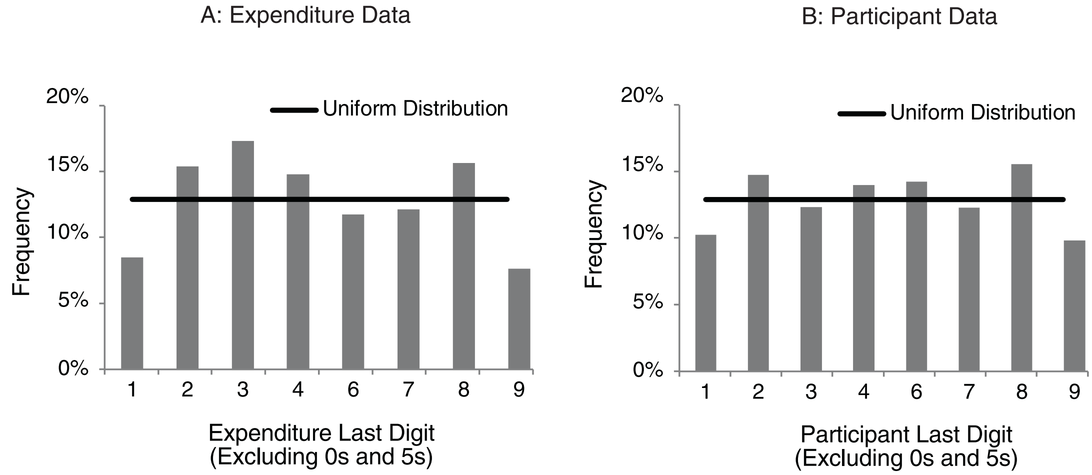

Figure 11. Last-digit expenditure and participant data against the uniform distribution.
Notes: This figure presents a last-digit test as compared to the uniform distribution, which is standard in the digit analysis literature. Expenditure data (Panel A) are statistically significant, with $p = 1.5 \times 10^{-9}$; $n = 851$. Participant data in Panel B shows preferences for the same (even) digits and are also statistically significant ($p = 7.0 \times 10^{-26}$; $n = 5850$). 0 s and 5 s are excluded from both tests due to rounding, which is tested separately in Figure 6.

Table 4
Significance of digit tests by district.

Fig	Digit Test	Mandera	Ijara	Wajir	Isiolo	Baringo	Garissa	Samburu	Marsabit	Moyale	Turkana	Tana	All Districts
2A	All Digit Places Beyond the First: Expenditure	**3.6E-14 846**	**2.6E-05 769**	**1.9E-06 1248**	0.0082 437	**7.3E-17 1352**	**2.8E-08 976**	0.020 848	**3.9E-04 449**	**1.5E-14 671**	0.40 907	**7.8E-04 868**	**3.9E-15 9371**
2B	All Digit Places Beyond the First: Participant	**9.0E-18 886**	**1.5E-10 765**	**6.5E-15 731**	**6.1E-11 478**	**2.1E-04 674**	**6.1E-18 858**	**2.3E-05 639**	0.25 527	0.033 736	**0.0037 591**	0.013 500	**5.5E-51 7385**
3	Padding Valuable Digit Places	**1.0E-05**	0.0054	**1.0E-05**	0.131	0.024	**0.0015**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**
4	Unpacking Rounded Numbers: Participant	**6.1E-21 453**	**1.1E-10 298**	**7.6E-11 433**	**4.4E-13 157**	0.0085 248	**5.9E-24 459**	**3.9E-05 179**	0.014 222	**0.0030 179**	**3.1E-05 205**	0.057 142	**2.5E-64 2975**
5	Election Year Effects: Expenditure	0.009	0.0098	**0.0001**	0.00605	0.0177	**0.00155**	**0.0001**	0.09075	**0.0001**	**0.0001**	0.01215	**0.001**
6	Rounding Digits: Expenditure	**8.7E-32**	**1.8E-06**	0.24	**5.3E-33**	1.0	0.86	1.0	**1.9E-38**	0.60	1.0	1.0	NA
7	Repeating Numbers: Expenditure	**2.6E-07**	0.036	0.98	**7.5E-04**	**4.0E-32**	1.0	1.0	1.0	1.0	1.0	0.98	NA
8	Sector Effects: Expenditure	**5.8E-21 373**	**3.2E-16 294**	**4.6E-14 338**	**5.5E-13 219**	**1.3E-16 424**	0.99 289	**1.2E-07 227**	**0.0035 211**	0.007 226	0.67 230	0.10 260	**5.8E-69 3091**
9	First-Digit: Expenditure Data	**1.4E-08 489**	**2.3E-13 386**	0.37 578	**5.5E-06 308**	**1.4E-09 488**	0.029 430	**5.7E-05 359**	0.011 293	**1.9E-12 319**	0.071 357	**0.0037 332**	0.089 4339
10	Digit Pairs: Participant	0.0070 238	**0.0029 176**	**4.9E-05 255**	1.0 125	**5.9E-04 251**	**1.2E-04 293**	0.35 166	**0.0025 126**	0.59 173	0.48 119	0.030 137	**2.4E-10 2059**
	Number of Significant Tests $p < 0.005$	8	7	7	6	6	6	6	5	5	4	3	

Notes: This table shows the *p*-value and sample size for each of 10-digit tests run on each of 11 districts. The tests were chosen to analyze different, non-overlapping aspects of the data. Given the large number of tests, a Bonferroni correction was used to establish 0.005 as the acceptable p – value for our tests. Statistically significant tests at the 0.005 level are indicated in bold. We tabulate the number of significant tests for each district in the bottom row.

ways. First, we show that the results of our tests are correlated with the World Bank forensic audit of the same project. Second, we validate our new all-digit-places approach with a secondary data source of self-reported financial data from businesses across the world and show that patterns of digit manipulation correlate with country-level measures of corruption.

6.1. *Establishing internal validity: Comparing digit analysis to the World Bank forensic audit*

The existence of an independent forensic audit for this World Bank project provides us with a unique opportunity to establish the internal validity of our new tests and to affirm the usefulness of digit analysis more broadly.

The measure of statistically significant digit tests presented in

Table 5
Digit tests by district compared to World bank INT forensic audit results.

District	Digit Tests (Number of Significant Out of 10)	INT Audit (Percent Suspected Fraudulent and Questionable Transactions)
Wajir	7	75
Isiolo	6	74
Samburu	6	68
Garissa	6	62
Tana	3	44
Mandera	8	Not Audited
Ijara	7	Not Audited
Baringo	6	Not Audited
Moyale	5	Not Audited
Marsabit	5	Not Audited
Turkana	4	Not Audited

Notes: This table shows the number of statistically significant digit analysis tests, out of 10, for each district in our data. Districts which fail greater levels of digit tests also have higher levels of suspected fraudulent and questionable transactions as measured by the World Bank forensic audit. A Pearson's correlation test of the 5 districts for which we have both digit tests, and the audit shows a correlation of 0.928, and a 95 % confidence interval of [0.255,0.995]. We reject the null hypothesis of no correlation at the 5 % significance level, with p = 0.023, t-statistic 4.33. The detailed results of each digit test are presented in Table 4. The source for the INT forensic audit data is (World Bank Integrity Vice Presidency, 2011).

Table 4 is correlated with the results of the World Bank's forensic audit, and we can reject a null correlation with $p < 0.05$. Table 5 compares the results of our digit analyses by district to the results of the World Bank forensic audit (World Bank Integrity Vice Presidency, 2011). 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 hypothesis for those same 4 districts in 6 to 7 of our 10 digit tests. The remaining district, Tana River, had the lowest levels of suspected fraud in the audit (44 percent); we reject the null on 3 of our 10 tests. A Pearson's correlation test of the 5 districts for which we have both digit tests and the World Bank audit shows a correlation of 0.928 and a 95 % confidence interval of [0.255, 0.995]. We reject the null hypothesis of no correlation at the 5 % significance level, with p = 0.0227.

We also find significant digit violations in all the unaudited districts we examine, 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 half (Mandera, Ijara, Baringo) have among the highest number of digit analysis violations (8, 7, and 6) in our sample. This underscores the potential gains of using digit analysis as a diagnostic for targeting costly auditing techniques to the areas of greatest suspicion.

6.2. Establishing external validity with the World Bank Enterprise survey

To externally validate our work, we run our all-digits test on a completely different dataset. We apply it to the World Bank Enterprise Survey, which asks businesses worldwide to report their financial positions. The data contains 179,063 observations from 154 countries. We examine 3 variables: past year sales, sales from 3 years ago, and total number of employees, each of which should conform to Benford's Law, and each of which can show behavioral limitations that arise if individuals make up values. To make the results compatible with our original analysis, we skip the first digit and omit the values 0 and 5. Because these data do not result in reimbursement, the padding test is not appropriate.

Appendix C Presents the results of this supplementary analysis. The level of deviation from Benford's Law of the reported data within a country is negatively correlated with the quality of governance, namely the control of corruption variable from the worldwide Governance indicators (Kaufmann & Kraay, 2020). This result is statistically significant, and it is robust to the use of any or all of the three variables; to limiting values from countries that have lower levels of rounding; and to the use of transparency International's CPI transparency rankings (Transparency International, 2023) in place of the WGI metric. Moreover, we show that the same preference for even digits appears among data from the worst-ranked control of corruption countries, and in countries neighboring Kenya (Uganda and Tanzania), while that pattern does not appear among the less corrupt countries

Our new analysis shows that the patterns detected by all-digit-places testing are not unique to our Kenyan context, nor to our specific dataset. Taken together with the simulations presented throughout the paper and in Appendix B, these findings confirm the broad applicability of digit analysis to financial statement and quantity reporting data worldwide.

7. Conclusion

Increased monitoring and oversight are important for development aid to reach its goal of helping the world's poor. Auditing development aid expenditures faces immense challenges, both in terms of the costs and difficulties of auditing on the ground in remote environments, as well as the missing incentives for development aid organizations to root out fraud or disclose negative findings.

In this paper, we present new methods specifically targeted to detect data tampering in development aid and other weak institutional contexts. Our statistical tests rely on expenditure and participant reports to find patterns consistent with profitable misreporting and attempts to evade detection. We demonstrate our methods on data from a World Bank project in Kenya. An independent forensic audit of the same project, as well as qualitative interviews and new simulations, confirm our digit analysis results, lending validity to the method and the substantive findings.

The exact battery of 10 tests that we use is not a turnkey system for digit analysis. Some characteristics of this dataset, such as the comparison of expenditure to beneficiary tests, are particular to these data, but also demonstrate the breadth of the approach. Our tests serve as an example of the power one can achieve with these techniques, though the specific tests used in other analyses will vary.

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 more challenging statistical exercise than ensuring that digits are uniformly distributed. It would require centralization across an organization, and matching of all supporting documentation, such as coordination of date-stamped receipts, cashbooks, vehicle logs, cancelled checks, and bank statements. Furthermore, everyone instructed to fabricate data would face an incentive to self-deal, which would undercut efforts to produce aggregate results consistent with Benford's Law. Such coordination would also expose leadership to a high risk of detection.

Digit analysis is especially beneficial in any circumstance where traditional forms of monitoring are challenging or expensive. It can be used in for-profit or other nonprofit settings, including as an additional layer of protection in traditional corporate accounting. We foresee the use of our method in a variety of new applications as well. Firms that invest in developing markets may choose to use this method to conduct their own form of monitoring. This method can also be used to test the authenticity of data supplied by governments in compliance with international environmental and financial agreements, or to verify pollution and labor data supplied for treaty compliance. In the modern world, where big data proliferates, stronger tools to analyze these data for signs of strategic and profitable manipulation will find increasing applicability.

CRediT authorship contribution statement

Jean Ensminger: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jetson Leder-Luis:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We particularly thank Avinash Dixit, Esther Duflo, Ben Gillen, Jonas Heese, Karla Hoff, Jonathan Katz, Pierre Liang, Ben Olken, Antonio Rangel, Eddie Riedl, Ethan Rouen and Robert Sherman for advice and support at critical junctures in 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, The World Bank, Massachusetts Institute of Technology, California Institute of Technology, New York University, Center for Global Development, Duke University, Oxford University, and the University of California (Irvine). We thank the California Institute of Technology for generous funding of this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.worlddev.2024.106858.

References

Adiguzel, F. S., Cansunar, A., & Corekcioglu, G. (2020). Truth or dare? Detecting systematic manipulation of COVID-19 statistics. *Journal of political institutions and political economy*.

Aguilar, M. A., Gill, J. B., & Pino, L. (2000). *Preventing Fraud and Corruption in World Bank Projects: A Guide for Staff*. Washington, DC: The World Bank.

Alt, J. E. (2018). *Twenty years of transparency research. Public Sector Economics 2018 – Fiscal openness: Transparency, participation and accountability in fiscal policies*. Zagreb: Institute of Public Finance, International Budget Partnership and Friedrich-Ebert-Stiftung.

Amiram, D., Bozanic, Z., & Rouen, E. (2015). Financial statement errors: Evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies, 20*, 1540–1593. https://doi.org/10.2139/ssrn.2374093

Andersen, J. J., Johannesen, N., & Rijkers, B. (2022). February). Elite Capture of Foreign Aid: Evidence from Offshore Bank Accounts. *Journal of Political Economy, 130*(2), 388–425.

Barabesi, L., Cerasa, A., Cerioli, A., & Perrotta, D. (2018). April). Goodness-of-Fit Testing for the Newcomb-Benford Law With Application to the Detection of Customs Fraud. *Journal of Business and Economic Statistics, 36*(2), 346–358.

Barney, B. J., & Schulzke, K. S. (2016). Moderating "cry wolf" events with excess MAD in Benford's Law research and practice. *Journal of forensic accounting research*, A66–A90.

Beber, B., & Scacco, A. (2012). What the numbers say: A digit-based test for election fraud. *Political Analysis, 20*, 211–234. https://doi.org/10.1093/pan/mps003

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76*, 169–217. https://doi.org/10.2307/1830482

Boland, P., & Hutchinson, K. (2000). Student selection of random digits. *The Statistician, 49*, 519–529.

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

Bredl, S., Winker, P., & Kötschau, K. (2012). A statistical approach to detect interviewer falsification of survey data. *Survey Methodology, 38*, 1–10.

Cerioli, A., Barabesi, L., Cerasa, A., Menegatti, M., & Perrotta, D. (2019, January 2). Newcomb–Benford law and the detection of frauds in international trade. *Proceedings of the National Academies of Science, 116*(1), 106-115.

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

Cheng, X., Palmon, D., Yang, Y., & Yin, C. (2022). January). Strategic Earnings Announcement Timing and Fraud Detection. *Journal of Business Ethics.*

Chin Hsien, H., & Lin, F. (2011). Applying digital analysis to detect fraud: An empirical analysis of US marine industry. *Applied Economics*, 135–140.

Cho, W. K., & Gaines, B. J. (2012). Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance. *The American Statistician, 61*, 218–223.

Claessens, S., Feijen, E., & Laeven, L. (2008). Political connections and preferential access to finance: The role of campaign contributions. *Journal of Financial Economics, 88*(3), 554–580.

Cuneo, M., Leder-Luis, J., & Vannutelli, S. (2023). Government Audits. *NBER Working Paper #30975*.

Dávid-Barrett, E., & Fazekas, M. (2020). Anti-corruption in aid-funded procurement: Is corruption reduced or merely displaced? *World Development*.

da Silva, C. G., & Carreira, P. M. (2013). Selecting Audit Samples Using Benford's Law. *AUDITING: A Journal of Practice & Theory*, 53–65.

Debreceny, R. S., & Gray, G. L. (2010). Data mining journal entries for fraud detection: An exploratory study. *International Journal of Accounting Information Systems*, 157–181.

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

Duflo, E., Greenstone, M., Pande, R., & Ryan, N. (2013). Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *The Quarterly Journal of Economics, 128*, 1499–1545. https://doi.org/10.1093/qje/qjt024

Dulleck, U., & Kerschbamer, R. (2006). March). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature, 44*(1), 5–42.

Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's Law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting, V*, 17–34.

Easterly, W. (2002). The cartel of good intentions: The problem of bureaucracy in foreign aid. *Journal of Policy Reform*, 223–250.

Easterly, W., & Williamson, C. R. (2011). Rhetoric versus Reality: The Best and Worst of Aid Agency Practices. *World Development, 39*(11).

Ensminger, J. (2017). *Corruption in community-driven development. A Kenyan case study with insights from Indonesia*. Bergen, Norway: Michelsen Institute, U-4 Anti-Corruption Resource Center.

Ferraz, C., & Finan, F. (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *Quarterly Journal of Economics.*

Gans-Morse, J., Borges, M., Makarin, A., Mannah-Blankson, T., Nickow, A., & Zhang, D. (2018). Reducing bureaucratic corruption: Interdisciplinary perspectives on what works. *World Development, 105*, 171–188.

Gibson, C. C., & Long, J. D. (2009, September). The presidential and parliamentary elections in Kenya, December 2007. *Electoral Studies, 28*(3), 497-502.

Hill, T. P. (1995). A statistical derivation of the significant-digit law. *Statistical Science, 10*, 354–363. https://doi.org/10.1214/ss/1177009869

Jansen, E. G. (2013). Don't rock the boat: Norway's difficulties in dealing with corruption in development aid. In T. Soreide, & A. Williams (Eds.), *Corruption, grabbing and development: Real world challenges*. Cheltenham, UK: Edward Elgar Publishing Limited.

Jones, C. (2020). *February 18)*. Financial times: The World Bank paper at the centre of a controversy.

Judge, G., & Schechter, L. (2009). Detecting problems in survey data using Benford's Law. *Journal of Human Resources, 44*, 1–24.

Karpoff, J. M., Lee, D. S., & Vendrzyk, V. P. (1999). Defense Procurement Fraud, Penalties, and Contractor Influence. *Journal of Political Economy*, 809–842.

Kaufmann, D., & Kraay, A. (2020). *Worldwide Governance Indicators*. World Bank.

Knepper, D., Lindblad, A. S., & Seifu, Y. (2016). Statistical Monitoring in Clinical Trials: Best Practices for Detecting Data Anomalies Suggestive of Fabrication or Misconduct. *Therapeutic Innovation & Regulatory. Science.*

Lamoreaux, P., Michas, P., & Schultz, W. (2015). Do Accounting and Audit Quality Affect World Bank Lending? *The Accounting Review, 90*(2), 703–738.

Mack, V., & Stoetzer, L. (2019). April). Election fraud, digit tests and how humans fabricate vote counts - An experimental approach. *Electoral Studies, 58*, 31–47.

Mebane, W. (2008). Election Forensics: The Second-Digit Benford's Law Test and Recent American Presidential Elections. In I. R. Alvarez, T. Hall, & S. Hyde (Eds.), *Election Fraud: Detecting and Deterring Electoral Manipulation*. Brookings Institution Press.

Michalski, T., & Stoltz, G. (2013). Do Countries Falsify Economic Data Strategically? Some Evidence That They Might. *The Review of Economics and Statistics, 95*, 591–616.

Nagi, M. H., Stockwell, E. G., & Snavley, L. M. (1973). Digit preference and avoidance in the age statistics of some recent African censuses: Some patterns and correlates. *International Statistical Review, 41*, 165–174. https://doi.org/10.2307/1402833

Nigrini, M. J. (2012). *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. Wiley & Sons.

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

Nirgini, M., & Miller, S. (2009). Data Diagnostics Using Second Order Tests of Benford's Law. *Auditing: A Journal of Practice & Theory, 28*(2), 305–324.

OECD. (2022, March 25). *Query Wizard for International Development Statistics.* Retrieved from OECD.org: https://stats.oecd.org/qwids/.

Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy, 115*, 200–249. https://doi.org/10.1086/517935

Olken, B. A., & Pande, R. (2012). Corruption in Developing Countries. *Annual Review of Economics*.

Rath, G. J. (1966). Randomization by humans. *The American Journal of Psychology, 79*, 97–103.

Reinikka, R., & Svensson, J. (2004). Local Capture: Evidence from a Central Government Transfer Program in Uganda. *Quarterly Journal of Economics, 119*(2), 679–705.

Republic of Kenya. (2006). *Arid Lands Resource Management Project (Phase II) Tana River District Progress Report 2003-2006.*.

Schräpler, J.-P. (2011). 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*, 685–718. https://doi.org/10.2139/ssrn.1562574

Stefanovic, M. (2018). *Former Manager of World Bank's External Investigations Unit, Integrity Vice Presidency*. Interviewer: J. Ensminger.

Svensson, J. (2005). Eight Questions about Corruption. *Journal of Economic Perspectives, 19*(3), 19–42.

Economist, T. (2020). *February 15)*. The Economist: The World Bank loses another chief economist.

The World Bank. (2019). *Net official development assistance and official aid received*. Retrieved from World Bank Data: https://data.worldbank.org/indicator/DT.ODA.ALLD.CD.

Toedter, K.-H. (2019). Benford's Law as an Indicator of Fraud in Economics. *German Economic Review*.

Transparency International. (2016). *Corruption in the Pharmaceutical Sector: Diagnosing the Challenges.*.

Transparency International. (2023). *Corruption Perceptions Index*. Retrieved from https://www.transparency.org/en/cpi/2022.

U.S. Agency for International Development Office of Inspector General. (2021). *Semiannual Report to Congress April 1, 2021-September 30, 2021.*.

U.S. Agency for International Development Office of Inspector General. (2021). *Semiannual Report to Congress October 1, 2020-March 31, 2021.*.

UN Economic and Social Council Economic Comission for Africa. (1986). Adjustment of Errors in the Reported Age-Sex Data from African Censuses . *Joint Conference of African Planners, Statisticians and Demographers.* Addis Ababa, Ethiopia.

World Bank. (2003). Project Appraisal Document on a Proposed Credit in the Amount of Sdr 43.6 Million (US $60m Equivalent) to the Republic ofKenya for the Arid Lands Resource Management Project Phase Two. *C.D. Eastern and Southern African Rural Development Operations, Africa. Region, 31*.

World Bank and OECD Development Assistance Committee. (2023). *Net official development assistance and official aid received (current US$)*. Retrieved from https://data.worldbank.org/indicator/DT.ODA.ALLD.CD.

World Bank Group Sanctions System. (2021). *Annual Report Fiscal Year 2021*. World Bank Group.

World Bank Integrity Vice Presidency. (2011). *Forensic Audit Report: Arid Lands Resource Management Project – Phase II – Redacted Report.* World Bank.

World Bank Integrity Vice Presidency and Internal Audit Department, Treasury, Government of Kenya. (2011). *Redacted Joint Review to Quantify Ineligible Expenditures for the Seven Districts and Headquarters of the Arid Lands Resource Management Program Phase II (ALRMP II) for FY07 & FY08*. Washington, DC: World Bank.