NBER WORKING PAPER SERIES

# DETECTING FRAUD IN DEVELOPMENT AID

Jean Ensminger
Jetson Leder-Luis

Working Paper 30768
http://www.nber.org/papers/w30768

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2022

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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Detecting Fraud in Development Aid
Jean Ensminger and Jetson Leder-Luis
NBER Working Paper No. 30768
December 2022
JEL No. C49,D73,H83,M42,O22

**ABSTRACT**

When organizations have limited accountability, antifraud measures, including auditing, often face barriers due to institutional resistance and practical difficulties on the ground. This is especially true in development aid, where aid organizations face incentives to suppress information about misappropriated funds and may operate with limited transparency. We develop new statistical tests to uncover strategic data manipulation consistent with fraud. These tests help identify falsified expense reports and facilitate monitoring in difficult-to-audit circumstances, relying only on mandated reporting of data. While the digits of naturally occurring data follow the Benford's Law distribution, humanly-produced data instead reflect behavioral biases and incentives to misreport. Our new tests improve upon existing Benford's Law tests by being sensitive to the value of digits reported, which distinguishes between intent to defraud and error, and by improving statistical power to allow for finer partitioning of the data.

We apply this method to a World Bank development project in Kenya. Our evidence is consistent with higher levels of fraud in harder to monitor sectors and in a Kenyan election year when graft also had political value. The results are validated by qualitative data and a forensic audit conducted by the World Bank. We produce simulations that demonstrate the superiority of our new tests to the standards in the field. Our tests are useful beyond development aid, including for monitoring corporate accounting and government expenditures.

Jean Ensminger
California Institute of Technology
HSS 228-77
Pasadena, CA 91125
jensming@hss.caltech.edu

Jetson Leder-Luis
Questrom School of Business
Boston University
595 Commonwealth Avenue
Boston, MA 02215
and NBER
jetson@bu.edu

A data appendix is available at http://www.nber.org/data-appendix/w30768

## 1. Introduction

Firms and governments around the world incur large financial losses due to fraud. Organizations rely on the financial reporting of their agents, who can exploit asymmetric information to divert financial resources. This asymmetry can arise between a firm's owners and their employees, between regulators and firms, or between the public and the bureaucrats who serve them. Abundant resources are devoted to closing these information gaps and improving the quality of reported data, including disclosure regulation, audits, monitoring, whistleblowing and, increasingly, tests of the data themselves to determine their quality. All reported data contain both information about the underlying true values, and also signals about the quality of reporting (Leuz & Wysocki, 2016). These signals provide an avenue through which fraud can be detected, which we explore.

Developing countries, and the aid organizations that serve them, face additional challenges when it comes to financial impropriety. The primary mechanisms for detecting and deterring corruption and fraud, such as auditing, adherence to accounting principles, and criminal and civil liability for corrupt individuals, require strong institutional support, as well as accountability, when rules or norms are violated. Aid organizations that serve these 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, combined with the weak institutional environments in developing countries, have made the application of traditional anti-fraud policy in the development aid space largely unsuccessful.

In this paper, we provide a partial solution to the challenges of monitoring expenditures and many other forms of data from countries and organizations that works even in weak institutional environments. Digit analysis analyzes the patterns of reported data to detect fraud, relying on the fact that humanly-generated data are different from naturally-occurring data. Humans face incentives to manipulate the data, as well as behavioral biases when producing data, while naturally-occurring data follow Benford's Law. We build upon earlier digit analysis work (such as Amiram *et al.* (2015), Barabesi *et al.* (2018), and Nigrini and Mittermaier (1997) ) to improve statistical power and present tests that better reveal suspected intent to defraud. In the developing world context, where enforcement of institutions mandating cooperation in audits may be weak, one advantage of this method is that it does not require the cooperation of potentially complicit suspects.

We apply these statistical tests to data from a World Bank development project in Kenya. The data contain details about development aid expenditures and numbers of beneficiaries served. Qualitative information based on hundreds of interviews points to high levels of graft from this 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 qualitative findings were confirmed by the forensic audit. The audit revealed that the Bank's financial controls, monitoring, and existing audit mechanisms were not capturing the extreme level of suspected fraud that existed. The World Bank audit flagged 66% of the district transactions as suspicious (49% as suspected fraudulent and 17% as questionable). One outcome of our method, which is the number of statistical tests failed per geographic district, is statistically significantly correlated with the level of suspected fraudulent and questionable transactions from the same districts examined in the forensic audit. The correspondence of our statistical tests with the forensic audit provides internal validity of the digit analysis method that, to the best of our knowledge, has not previously been reported in the literature.

Naturally-occurring data and humanly-produced data are different along several dimensions. Humans face behavioral limitations in producing numbers (Chapanis, 1995), and 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 which 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.

We advance the existing digit analysis and Benford's Law literatures in several ways. First, we expand the statistical power of Benford's Law goodness of fit testing by considering all digit places in one test, rather than just one or two digit places, as is the norm in previous literature. By improving statistical power, we allow for additional disaggregation and triangulation of data categories. Second, the existing Benford's Law literature has focused on aberrant patterns, but has limited capacity to distinguish between strategic misreporting, which seeks to gain profit for the fraudster and subvert detection, and benign misreporting or error. This issue is driven by the fact that Benford's Law predicts digit distributions from the front of the number (e.g., first digit, second digit), irrespective of the number's value (i.e., one thousand versus one hundred thousand). This means that basic tests of conformance to Benford's Law are not sensitive to the value of the digit being manipulated. One of our tests considers the value of the number and allows us to distinguish patterns consistent with profitable misreporting. This moves the evidence closer to establishing an intent to defraud. We supplement our 2 new tests with the results of 8 other tests, including tests from the existing literature, that capture economic and political incentives to steal, as well as the behavioral patterns that arise when humans fabricate data. This allows us to compare the digit test results to the results of the World Bank's forensic audit of the same project. Third, we validate our statistical tests using extensive simulations to show that our tests can successfully detect misreporting in a way that is not specific to this case study, and also that the patterns we uncover are not driven by benign factors such as underlying prices.

Our work reveals some important substantive findings that underscore its advantage relative to existing Benford's Law tests. 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 being 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 (see, (Olken B. A., 2007)).

Our digit-based method for uncovering fraud is complementary to other popular forms of anti-fraud machine learning, which have focused on reported values such as the debt-to-equity ratio, or institutional details like the presence of a Big 4 auditor (Perols, 2011). Therefore, we anticipate that this method will have applicability both for measuring earnings fidelity as well as for detecting fraud in U.S. public firms and in developing world investments. Indeed, the SEC has recently warned that investors in emerging markets face risk due to limited and unreliable financial reporting (U.S. Securities and Exchange Commission, 2020). Our method can be used for ongoing monitoring to achieve early detection of irregularities, and to assist audits by guiding sample selection for deeper investigation.

Our work relates to a large body of literature that has addressed audit quality, the organizational economics of fraud, and the incentives of auditors. Auditing faces the challenge that it is costly, and also that auditors are often employed by the very people that they monitor, generating conflicting incentives to report suspected impropriety. Goldman and Barlev (1974) discuss threats to auditor independence and the conflicts of interest they face, particularly from management that controls their employment and wants a favorable report. Their paper came early in a large literature on auditor independence, which has since paid much attention to the financial and public sectors, but less to international aid flows, where the problem is arguably more dire. Our paper also sheds some light on the problem of audit quality in weak institutional environments. Krishnan *et al*. (2006) show that misreporting among nonprofits is driven by managerial incentives and disciplined by the use of outside accountants. In a recent paper on auditor and client relationships, Cook *et al.* (2020) demonstrate that the reputations of auditors tend to match the reputations of the clients they serve.

Lamoreaux *et al*. (2015) find that World Bank development aid loans are higher for countries with better accounting quality, but accounting issues 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. In other related work, Duflo *et al*. (2013) provide an example of auditor capture in the developing world and show that monitoring of monitors is an effective way to combat fraud. These papers establish the need for new tools to address such challenging environments.

Digit analysis and Benford's Law have generated a long literature of statistical methods. We discuss this literature fully and its relationship to our analysis in Section 3.3.

## 2. 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 investments in projects to promote growth and development in poor countries. They do this through bilateral aid, such as USAID, and through multilateral aid such as the World Bank and the European Union. The U.S. alone spends about $45 billion per year on these endeavors. In short, aid is a major worldwide industry, but with vastly different oversight incentives and institutions than the for-profit sector.

Development aid faces a fundamental challenge in fighting graft: aid money flows to the poorest parts of the world, which have the weakest institutional environments, greatly increasing the risk of embezzlement. 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 (Kaufmann & Kraay, 2020) (The World Bank, 2019). 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, 92% of aid flows to countries with a below-mean corruption control measure, indicating the scale of the threat that aid dollars face.

[FIGURE 1 HERE]

Monitoring mechanisms such as auditing are used to control and detect fraud in a variety of contexts (Anderson, Francis, & Stokes, 1993), and empirical evidence has shown that increased auditing is effective at eliminating graft in development aid (Olken B. A., 2007). The World Bank has historically relied on internal investigations and monitoring tools such as whistleblower hotlines and internal audits as its primary anti-fraud mechanisms (Aguilar, Gill, & Pino, 2000). However, the usefulness of these tools relies on the ability and willingness of development aid staff or beneficiaries 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 building, goods and equipment, services such as health care or child education, and trainings for beneficiaries to improve their human capital on 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 quantity surveyor (Olken B. A., 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, as the good produced is intangible human capital, attested to by beneficiaries who may be difficult to find, and can face retaliation from the project for negative statements to outside monitors.

Development organizations face conflicts of interest. Development organizations often depend upon the field-supervision of outside experts who are typically chosen by the staff member overseeing the project. Their employment on future missions may depend upon reports favorable to the project. Routine financial management and auditing is usually handled internally by understaffed departments. The World Bank Integrity Vice Presidency (INT) is responsible for the Bank’s fraud investigations, and similar responsibilities are held by the Office of the Inspector General for USAID (OIG-USAID). 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, 2021) (U.S. Agency for International Development Office of Inspector General, 2021). 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 has taken place for any World Bank project before or since this one. This is the only such audit on the World Bank Internal Investigations website (World Bank Integrity Vice Presidency, 2011).

A primary factor in the low rates of auditing in developing contexts is the lack of incentives to monitor, and the direct incentives not to disclose 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, and do so 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 principal, developed countries, must rely on the agent, the development aid organization, both to *administer* the aid, and also to *monitor their own performance*. And yet, 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.

Recent literature addresses the issue of the subversion of aid at a macro level and shows the incentives of development aid organizations to suppress these facts. Andersen, Johannesen and Rijkers (2022) show 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. However, the author of that study, Bob Rijkers, is an employee of the World Bank, and his attempts to publish this 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 resigning unexpectedly and shortly following this incident (Jones, 2020), (The Economist, 2020). This case underscores both the magnitude of fraud in development aid as well as the missing incentives for development aid organizations to effectively monitor themselves and disclose their negative findings.

Qualitative data also point to high levels of graft and low levels of anti-fraud measures in development aid projects. Appendix B presents data from interviews concerning the World Bank Arid Lands project in Kenya, which is the project examined in this paper. 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, for which Jansen was the program officer. Jansen identifies the lack of external monitoring, and attempts to suppress internal monitoring, that are consistent with our qualitative evidence.

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 data that they already collect. This disclosure could be mandated by donor nations who fund development aid organizations. Digit analysis does not require the cooperation of potentially complicit subjects and can be used to detect signals of fraud and to guide deeper investigations. By mandating data transparency, rather than pushing aid organizations to audit, donors can more easily ensure compliance. Digit analysis can 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 organizations.

## 3. Research Method: Theory and Motivation

Given the challenges of monitoring development aid, digit analysis provides a method to detect fraud that requires only that data be made available. Here, we present 2 new statistical tests that provide a method for examining falsified data in development aid and beyond. We then utilize these 2 new tests in applications to the project dataset that take full advantage of their power. We complement these 4 tests with 6 additional tests (see Appendix A) that further display the incentives and behavioral limitations of those who fabricate data.

We motivate our statistical testing with a theoretical framework for the incentives of those who are tasked with producing expenditure reports. Those who report, typically bureaucrats, face a decision either to accurately report spending or to fabricate such data. The statistical properties of the observed data result from this decision, and this theoretical framework provides predictions of 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, as they accurately reflect the data without human interference.

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

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

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 4$^{th}$ 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 (Amiram, Bozanic, & Rouen, 2015; Diekmann, 2007; Hill, 1995) (Nigrini & Mittermaier, 1997).

[Table 1 here]

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, 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 C 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 such as those represented in our data, 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 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 behavior 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 arrangements 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.[1] 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).

It is possible that specific digit preferences are culturally influenced, in which case it is instructive to have a culturally representative baseline for comparison. Evidence of specific digit

[1] 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.

preferences from Africa comes from an examination of African census data. A phenomenon known as age heaping occurs when people are approximating 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 Comission for Africa, 1986). These same digit patterns occur in our data; both 0 and 5 are so heavily overrepresented that we omit them in most of our analyses and analyze only digits 1-4 and 6-9. Nevertheless, we can rule out the idea that the patterns present in our data are the result of *legitimate* digit preferences for underlying price. Appendix C 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.

*3.3 Digit Analysis Literature*

Our method builds upon the widespread but previously underpowered use of digit analysis for the detection of anomalies and fraud, and we expect that our method can improve upon existing applications of digit analysis. Digit analysis has been used in accounting to measure financial statement errors (Amiram, Bozanic, & Rouen, 2015), as well as in forensic auditing, where it is used for targeting deeper investigation (Nigrini & Mittermaier, 1997; Durtschi, Hillison, & Pacini, 2004). However, these applications rely on one- or two-digit-place comparisons, often in the first, second, or last digit, which limits statistical power and can run into sample size concerns.

A set of literature has described more advanced statistical procedures to perform Benford's Law testing. Nigrini and Miller (2009) consider a second-order test of conformance to Benford's Law, which considers the difference between ranked values in a dataset, these differences themselves being tested for Benford conformance. Da Silva and Carreira (2013) use Benford's law to find specific subsets of the data with the greatest nonconformance, to assist auditors with further investigation. Barabesi *et al*. (2018) apply digit analysis tests to detecting customs fraud using a sequential tree-structured testing procedure called serial gatekeeping, 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 since values in international trade data may not be Benford conforming. In each of these papers, the authors tune tests for conformance to the Benford distribution to improve power or target the test or the sample based solely on the Benford's Law distribution. Our work complements these studies by also considering the political incentives to divert funds, the specific behavioral limitations of those fabricating data, and the financial incentives to pad values in valuable digit places. Our test of all digit places removes the need for sequential testing, incorporating all the statistical power of Benford's law into one easy-to-use test that itself can be disaggregated to find appropriate subsamples.

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). Other areas where digit analysis has been successfully used include in the detection of IMF data manipulation (Michalski & Stoltz, 2013), campaign finance fraud (Cho & Gaines, 2012), scientific data fabrication (Diekmann, 2007), and enumerator integrity during survey research (Bredl, Winker, & Kötschau, 2012; Judge & Schechter, 2009; Schräpler, 2011). The ever-increasing value of data leads to greater incentives to manipulate that data and has led researchers to use digit analysis in a variety of new settings.

Our analysis also contributes to the accounting and economics literature focused on monitoring, anomaly detection, and the measurement of data quality. Du *et al.* (2020) measure the fidelity of firms' reported earnings using a hidden Markov model and show that this can predict external indicators of bad accounting, specifically Security and Exchange Commission comment letters and earnings restatements. Perols *et al.* (2017) provide another method for fraud detection using data analytic methods. These authors use the reported values from accounting statements of known fraudulent firms to classify other firms as suspected fraudulent. Our method complements these existing studies; while these papers rely on the *values* of self-reported data, our method relies instead on the *patterns* of such data. As such, we expect that our measurement could be incorporated into broader models of earnings fidelity in future accounting studies. With its focus on pattern analysis, our work is similar to Purda and Skillicorn (2015), who analyze the text of annual and interim corporate reports and show that language patterns can be used for statistical detection of fraud.

## 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. Our digit analysis is confined to the 11 arid districts, as these districts were the most homogeneous across ecological, economic, and demographic measures. 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.

The expenditure and participant data used in these analyses were extracted from quarterly electronic project reports produced by each of the 11 districts. 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.

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. Additional details about these data, as well as qualitative details about the nature of corruption in the Arid Lands project, are presented in Appendix B.

*4.2 Forensic Audit Data*

In 2009, following an external complaint, the World Bank's Integrity Vice Presidency (INT) began a broad forensic audit of the Arid Lands project that lasted 2 years and culminated in a public report (World Bank Integrity Vice Presidency, 2011).[2] Auditors sampled 2 years' worth of receipts for 7 districts, 5 of which were arid districts examined in this analysis. They examined 28,000 transactions. 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. Two of our tests are new, and two more build upon our new tests to show specific examples of data manipulation. The remaining 6 tests are variations on existing tests in the literature and are presented in Appendix A. 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

---

[2] 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.

use a Bonferroni correction: we divide our desired significance level (.05) by the number of tests (10) and set a significant level of $p = .005$, 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 Appendix A.

*5.1. Increasing Statistical Power: All Digit Places Beyond the First*

A simple, 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, resulting in low $n$. Furthermore, testing individual digit places results in multiple-hypothesis testing issues, which a two-way chi square test 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 in Appendix A.3.

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 in a test for excess rounding (see Appendix A.1). For each digit place ($2^{nd}$ digit, $3^{rd}$ 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). The data 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 non-conformance to the expected Benford distribution ($p = 3.9 \times 10^{-15}$). Strikingly, these same digit patterns appear 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 with the humanly generated African census pattern described earlier.

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.

[Figure 2 Here]

The lack of conformance to the expected distribution, consistency with known humanly-generated data from African census studies, and similar patterns across both expenditure and participant data, are strong indicators that these data have been tampered with. Importantly, this test of multiple digit places subsumes a test of last digits alone, which we present as a robustness check in Appendix A.5.

5.1.1 Simulations on the Power of All-Digit-Place Testing

Our new single test explores patterns in all digit places simultaneously, rather than multiple tests of different digit places, which greatly improves statistical power. An extensive literature on forensic auditing and the use of digit analysis have promoted the use of single-digit-place tests to find evidence of fraud, or to select samples of data for additional review or auditing. These tests focus on the statistics comparing a single digit, such as the first digit, second digit, or last digit, to Benford's Law (see, e.g. Nigrini and Mittermaier (1997) and (Beber

& Scacco, 2012)). Here, we use a simulation to exhibit the relative power of this test as opposed to single-digit-place testing.

Our simulation proceeds as follows. We generate Benford-conforming data between 4 and 8 digits long, (i.e., between 1,000 and 99,999,999), with each of 6 simulated districts having 1,000 observations of data. We simulate 3 "bad" districts in the data, districts A, B, and C, which each have a preference for 2 digits chosen independently. For example, district A might prefer 3 and 7, while district B might prefer 2 and 5. For each bad district, each observation is originally generated as conformant to Benford's Law, but there is a 20% chance that they manipulate the data by replacing a digit in that observation with their preferred digit. There are also 3 "good" districts, D, E, and F, which produce Benford-conforming data with no digit preferences. An ideal test would be able to distinguish good districts from bad districts and successfully flag districts A, B, and C for further review, while not flagging districts D, E, and F.

To test the current standard in the literature, we first consider tests of first digits, second digits, third digits, and last digits in each of these districts. Given a desired significance threshold of $p < 0.05$, we must correct for multiple testing by dividing by the number of tests (4) and achieve a significance threshold of 0.0125.

Table 2, Panel A presents the results of these tests. The sample size for each test is 1,000. Panel A shows the issues with single-digit testing. In the first digit, no districts are statistically significant, and in the second digit, only district A stands out. No districts are statistically significant in the third digit. Pooling data from different districts similarly fails to detect aberrant patterns using these tests. In the last digit, District F is inappropriately flagged as suspicious. Raising the statistical significance threshold back to 5%, that is, ignoring the Bonferroni correction for multiple tests, does not fix this issue; indeed, it would flag district E as suspicious in the last digit as well. District B fails no tests despite being (statistically) equally as manipulated as districts A and C.

[Table 2 Here]

Importantly, last digits here are tested against the uniform distribution, as is promoted by the literature (see Beber *et al* (2012)). District F, which has no manipulation, fails this test. The uniform distribution is generally appropriate for last digits, but last digits may have slight tendencies towards Benford's law when they are also part of short numbers. As seen in Table 1, in a 3-digit number, the last digits are third digits, which are not uniformly distributed. Here, the smallest number is 1,000, so the last digit place is the $4^{th}$ digit place for these numbers. Therefore, the last-digit test here produces a false positive, and indeed is marginally significant ($p < 0.10$) for every district.

Table 2, Panel B presents an alternative testing regime, where we consider our new test of all digit places by district. This is a single test, and the appropriate statistical significance threshold is 5%. The sample sizes vary slightly because exact values are simulated, so some districts have more digits than others due to the random length of numbers. The three manipulated districts fail this test (A, B, and C), as they should, and none of the unmanipulated districts do (D, E, and F), as they should.

Appendix C.2. extends the results of this analysis to different sample sizes ($n$) and different rates of manipulation ($p$) using the same setup of 6 districts. These extended simulations show that the all-digit-places test outperforms many single-digit-place tests, with a higher true-positive rate and a lower false-positive rate among a range of sample sizes and manipulation rates.

This simulation shows how the all-digit-places test substantially outperforms single-digit testing along many dimensions. Signals of fraud may be present in different digit places, but individual-digit-place tests fail to combine these signals in statistically powerful ways. When performing single-digit testing, each test must be compared to a significance threshold, but each test fails to incorporate corroborating information in different digit places. Our new multiple-digit-places test exactly solves this issue, improving the sample size and power of each test, and picking up digit preferences that are observable when a reporter exhibits them over different digit places.

*5.2. Strategic Intent: Padding Valuable Digit Places*

The first test demonstrates that the data do not conform to Benford's Law but does not demonstrate the directionality of how people are manipulating the digits. 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 fraud manipulation and accidental data error. While there may be a strong correlation between firms and individuals whose paperwork is sometimes incomplete or missing, 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 deducing intent to defraud.

As discussed in Section 3.2, bureaucrats falsifying data can be expected to inflate values in order 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 ($1^{st}$ digit, $2^{nd}$ digit). However, the value of a digit depends on the digit's position from the right (e.g., 1s, 10s, 100s place), and this value determines the incentive to manipulate a digit. Therefore, basic tests of conformance to Benford's law are not sensitive to the value of the digit being manipulated.

To overcome this limitation, we compute the expected mean under Benford's Law by digit place *from the right* (10s, 100s), using the length of the numbers in our dataset to match left-aligned digit places and right-aligned digit places. 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. We then perform a Monte Carlo simulation of 100,000 Benford-distributed digits in each digit place, compare the difference-of-means statistic of the project data to the simulated data, and find the probability of observing our results under the Benford distribution. Appendix A.6 contains technical details of this process.

Figure 3 shows the padding tests among both World Bank and simulated data against the Benford expected distribution. The 0 line indicates the Benford mean; anything above the line represents an overuse of high digits, and anything below the line represents an underuse. The World Bank project data (Panel A) in the 10,000s 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.

[Figure 3 here]

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 10,000s to the 1,000s, 100s, 10s, and 1s 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 numbers 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. Behavioral Limitations: 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. The authors of the annual district reports also had reason to expect that participant data would not be as carefully scrutinized as expenditure data. First, the impact of participants on expenditures was obscured because it was only one component of the full costs of a single training exercise. Second, training exercises in remote villages are very difficult to verify because their final product is knowledge, which leaves limited 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, undertaking low-effort fabrication and reporting a round total number of participants to meet that target. This total number of participants was then split into males and females, as was required for reporting. 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 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 2s and 8s and low incidence of 1s and 9s when the gender specific data come from a pooled number that ends in 0 (Figure 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 are 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 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 shows the power of the all-digit-places test for disaggregating data to pick up specific patterns consistent with the behavioral limitations of those producing data. Analyses of single digit places struggle with disaggregation due to low sample sizes. Analyzing multiple digit places solves this issue.

[Figure 4 here]

*5.4. Election Year Effects*

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)(Fleming, Hermanson, Kranacher, & Riley Jr., 2016). Our next test makes use of our padding test to examine the timing of padded digits while simultaneously providing evidence consistent with intent to defraud.

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.

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 a highly controversial presidential election year that led to extreme violence (Gibson & Long, 2009).

[Figure 5 here]

This example further demonstrates the power of the padding test. It can be deployed to examine data fabrication in conjunction with time-based analysis. This test is sensitive to the patterns of humans fabricating numbers profitably, and is powerful enough to pick up signals even when disaggregating data.

*5.5 Other Tests*

Appendix A presents the results of 6 other tests that also exhibit the behavioral limitations and economic incentives expected from fabricated data. These tests are each motivated by 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 as a substring in the number 422,347), and last digits. For the study of rounding and repeats, we compare districts to each other, relying on the fact that patterns ought to be similar across districts, even when the appropriate baseline level of rounded or repeated data is unknown. These tests all corroborate that the World Bank data are highly manipulated and allow us to graphically and statistically examine different signals of this behavior.

*5.6 Summary of Tests*

Table 3 compiles the results of all 10 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 almost all overlap and pinpoint different aspects of data manipulation. In the bottom row, we sum the number of failed tests by district, which averages 5.7 out of 10, and ranges from 3 to 8.

[Table 3 here]

## 6. Establishing Validity: Comparing Digit Analysis to The World Bank Forensic Audit and to Qualitative Data from the Field

The existence of both an independent forensic audit for this World Bank project and qualitative data from the field 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.

*6.1 The World Bank Forensic Audit*

The measure of failed digit tests presented in Table 3 is statistically significantly correlated with the results of the World Bank's forensic audit. Table 4 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 lower levels of suspected fraud in the audit than the other districts (44 percent), and we rejected 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 $p = 0.0227$. The World Bank's forensic audit confirms the findings from our digit analysis tests.

[Table 4 here]

We also find significant digit violations in all of 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. 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 realities of auditing on the ground in challenging 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. These methods rely only on mandated reporting of data, something that most organizations already require. These methods require minimal cooperation from those who may be implicated in the fraud and who may have an incentive to impede an audit. We demonstrate our methods on data from a World Bank project in Kenya. Our statistical tests rely on expenditure reports to find patterns consistent with profitable misreporting and attempts to evade detection. An independent forensic audit of the same project, as well as qualitative interviews and new simulations, correlate with our digit analysis results, lending internal validity to the method and the substantive findings. This approach can reduce the information gap that enables misreporting between a principal and an agent.

Our method involves new statistical tests that will be broadly applicable to the study of misreported data beyond the context of development aid or developing countries in general. Our new test of padding valuable digits relates aberrant digit patterns to the monetary value of the digit place and uncovers patterns consistent with profitable deviations as well as attempts to evade detection. This test has the potential to differentiate intent to defraud from benign error. It can also be applied to test for evidence of intent to defraud with the timing of expenditure reporting. Another of our new tests, employing Benford's Law to analyze multiple digit places simultaneously, provides a statistically powerful test applicable to even relatively small datasets. The ability to work on smaller sample sizes allows more multi-dimensional analyses, such as our comparisons across districts, years, and sectors.

We extensively validate our methods using simulated data. First, we show that financial data should follow Benford's Law, even when underlying prices might be contaminated with digit preferences. This rules out the hypothesis that the patterns in our dataset are benign reflections of underlying patterns of prices set by vendors. Then, we show that our tests are powerful as compared to the standard in previous literature. Our simulations also show that our new simultaneous test of multiple digits outperforms single-digit place tests, allowing the user to disaggregate data and pinpoint fraud. Furthermore, our new test of padding is validated by data simulated from Benford's Law, which show no such pattern.

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. The exact set of tests that can be performed on other datasets depends on both the incentives for manipulation in that dataset, as well as the specifics of the attribute data that are available. Our tests serve as an example of the power one can achieve with these techniques and allow us to validate new and existing digit analysis tests with the overall results of the World Bank forensic audit.

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 is 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, each individual 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. It is not possible both to pad values, consistent with theft, and to conform to Benford's Law.

The substantive findings of this project attest to the need for better measures and identification of fraud in the developing world and under weak institutional environments everywhere. The World Bank's forensic auditors determined that 66% of the district transactions they examined were suspected fraudulent or questionable. Similarly, the districts we examined failed between 3 and 8 of the 10 digit tests that capture different dimensions of data manipulation. We demonstrate that more suspicious patterns emerge in a presidential election year, consistent with allegations that World Bank funds were illegally diverted to fund political campaigns. Our method could have been used to conduct real-time monitoring of this project to reduce potential fraud and to target the forensic audit of this project on the worst offending districts, 3 of which were missed in the World Bank audit sample.

Digit analysis is especially beneficial in any circumstance where traditional forms of monitoring are challenging or expensive, and it is applicable to any setting where individuals have incentives to fabricate data. 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.

## Bibliography


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.

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. doi: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.

Anderson, D., Francis, J., & Stokes, D. (1993). Auditing, directorships and the demand for monitoring. *Journal of Accounting and Public Policy, 12*(4), 353-375.

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.

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

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

Berkman, S. (2008). *The World Bank and the Gods of Lending* . Sterling, VA, USA: Kumarian Press.

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.

Chavis, L. (2010, November). Decentralizing development. Allocating public goods via competition.”. *Journal of Development Economics 93 (2): 264-74., 93*(2), 264-274.

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

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.

Cook, J., Kowaleski, Z., Minnis, M., Sutherland, A., & Zehms, K. (2020, August). Auditors are known by the companies they keep. *Journal of Accounting and Economics, 70*(1).

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

Debowski, L. (2003). Benford's Law Number Generator. Polish Academy of Sciences, Institute of Computer Sciences.

Deckert, J., Myagkov, M., & Ordeshook, P. (2011). Benford's Law and the detection of election fraud. *Political Analysis, 19*, 245-268. doi:10.1093/pan/mpr014

DeFond, M., & Zhang, J. (2014, November-December). A review of archival auditing research. *Journal of Accounting and Economics*, 275-326.

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

Du, K., Huddart, S., Xue, L., & Zhang, Y. (2020, April-May). Using a Hidden Markov Model to Measure Earnings Quality. *Journal of Accounting and Economics, 69*(2-3).

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. doi:10.1093/qje/qjt024

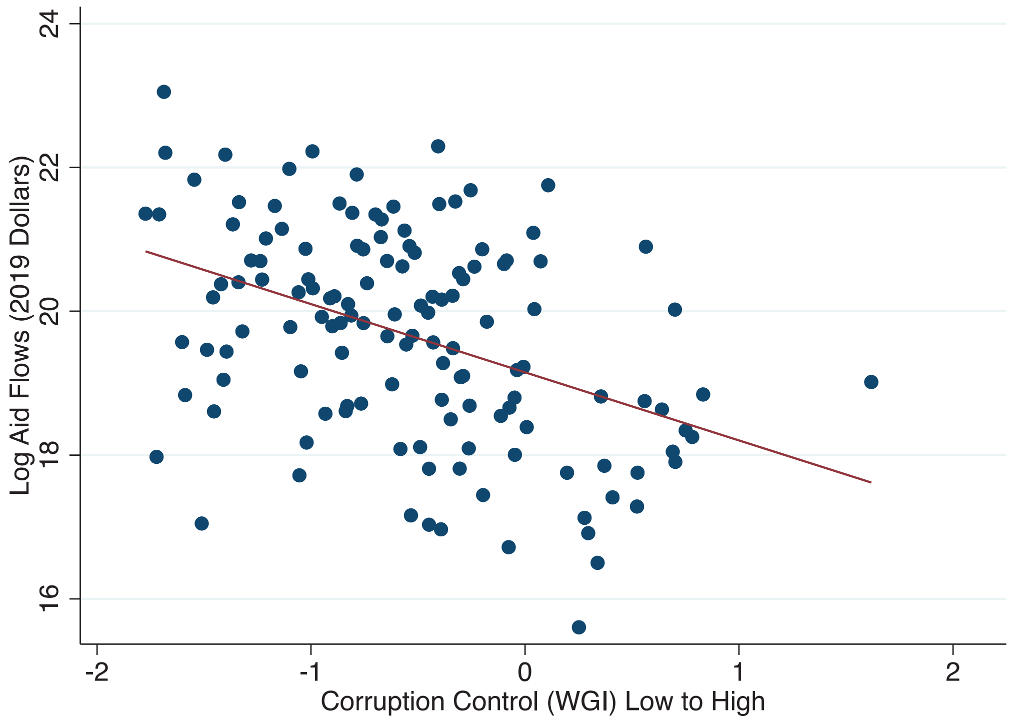
Dulleck, U., & Kerschbamer, R. (2006, March). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature, 44*(1), 5–42.

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.
Ensminger, J. (2017). *Corruption in Community Driven Development: A Kenyan Case Study with Insights from Indonesia.* U-4 Anti-corruption Resource Centre. Bergen, Norway: Chr. Michelsen Institute.
Fang, H., & Gong, Q. (2017). Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked. *American Economic Review , 107*(2), 562–591.
Fleming, A. S., Hermanson, D. R., Kranacher, M.-J., & Riley Jr., R. A. (2016). Financial Reporting Fraud: Public and Private Companies. *Journal of Forensic Accounting Research , 1*(1).
Fortune. (2020, March 25 2022). *Global 500*. Retrieved from Fortune.
Gibson, C. C., & Long, J. D. (2009, September). The presidential and parliamentary elections in Kenya, December 2007. *Electoral Studies, 28*(3), 497-502.
Goldman, A., & Barlev , B. (1974, Oct). The Auditor-Firm Conflict of Interests: Its Implications for Independence . *The Accounting Review , 49*(4).
Guggenheim, S. (2006). Crises and Contradictions: Understanding the Origins of a Community Development Project in Indonesia. In A. Bebbington, M. Woolcock, S. Guggenheim, & E. Olson, *The Search for Empowerment: Social Capital as Idea and Practice at the World Bank* (pp. 111-44). Bloomfield, CT: Kumarian Press.
Guggenheim, S., & Wong, S. (2005). Community-Driven Development: Decentralization's accountability challenge. In *East Asia decentralizes: Making local government work* (pp. 253-67). World Bank: Washington, DC.
Hill, T. P. (1995). A statistical derivation of the significant-digit law. *Statistical Science, 10*, 354-363. doi: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, *Corruption, grabbing and development: Real world challenges.* Cheltenham, UK: Edward Elgar Publishing Limited.
Jones, C. (2020, February 18). The World Bank paper at the centre of a controversy. *Financial times*.
Judge, G., & Schechter, L. (2009). Detecting problems in survey data using Benford's Law. *Journal of Human Resources, 44*, 1-24.
Kaufmann, D., & Kraay, A. (2020). *Worldwide Governance Indicators.* World Bank.
Krishnan, R., Yetman, M., & Yetman, R. (2006, March). Expense Misreporting in Nonprofit Organizations. *The Accounting Review, 81*(2), 399-420.
Lamoreaux, P., Michas, P., & Schultz, W. (2015). Do Accounting and Audit Quality Affect World Bank Lending? . *The Accounting Review, 90*(2), 703–738.
Leuz, C., & Wysocki, P. (2016). The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research . *Journal of Accounting Research, 54*(2), 525-622.
Mack, V., & Stoetzer, L. (2019, April). Election fraud, digit tests and how humans fabricate vote counts - An experimental approach. *Electoral Studies, 58*, 31-47.
Mansuri, G., & Rao, V. (2013). *Localizing Development: Does Participation Work?* Washington, DC: World Bank.
Mebane, W. (2008). Election Forensics: The Second-Digit Benford's Law Test and Recent American Presidential Elections. In R. Alvarez, T. Hall, & S. Hyde, *Election Fraud: Detecting and Deterring Electoral Manipulation.* Brookings Institution Press. Retrieved from http://www.jstor.org/stable/10.7864/j.ctt6wpf99
Mebane, W. (2011). Comment on "Benford's Law and the Detection of Election Fraud". *Political Analysis*(19), 269-272.
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. doi:10.2307/1402833
Nigrini, M. (2012). *Benford's Law.* Hoboken, New, Jersey: John Wiley & Sons, Inc.
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. doi:10.1086/517935

Olken, B. A. (2009). Corruption Perceptions vs. Corruption Reality. *Journal of Public Economics, 93*(7-8), 950-64.
Perols, J. (2011, May). Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. *Auditing: A Journal of Practice & Theory, 30*(2), 19-50.
Perols, J., Bowen, R., Zimmermann, C., & Samba, B. (2017). Finding Needles in a Haystack: Using Data Analytics to Improve Fraud Prediction . *The Accounting Review, 92*(2), 221-245.
Purda, L., & Skillicorn, D. (2015, Fall). Accounting Variables, Deception, and a Bag of Words: Assessing the Tools of Fraud Detection. *Contemporary Accounting Research, 32*(3), 1193-1223.
Rath, G. J. (1966). Randomization by humans. *The American Journal of Psychology, 79*, 97-103.
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. doi:10.2139/ssrn.1562574
Stefanovic, M. (2018). Former Manager of World Bank's External Investigations Unit, Integrity Vice Presidency. (J. Ensminger, Interviewer)
The Economist. (2020, February 15). The World Bank loses another chief economist. *The 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
Transparency International. (2009). *Corruption Perceptions Index*. Retrieved August 10, 2020, from https://www.transparency.org/en/cpi/2009#
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.*
U.S. Securities and Exchange Commission. (2020, April 21). *Emerging Market Investments Entail Significant Disclosure, Financial Reporting and Other Risks; Remedies are Limited.* Retrieved July 30, 2020, from SEC Public Statements: https://www.sec.gov/news/public-statement/emerging-market-investments-disclosure-reporting
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.
United Nations Inter-agency Task Force on Financing for Development. (2021). *Data update to the 2021 Financing for Sustainable Development Report, following the 13 April release of 2020 ODA data.* New York.
Wong, S. (2003). *Indonesia Kecamatan Development Program. Building a monitoring and evaluation system for a large-scale community-driven development program.* Washington, DC: World Bank.
Wong, S., & Guggenehim, S. (2018). *Community-Driven Development: Myths and Realities.* World Bank Group: Social, Urban, Rural and Resilience Global Practice.
Woodhouse, A. (2002). *Village corruption in Indonesia: Fighting corruption in the World Bank's Kecamatan Development Program.".* Washington, D.C.: World Bank.
Woodhouse, A. (2012). *Governance Review of PNPM Rural, Community level analysis.* Jakarta, Indonesia: World Bank Indonesia.
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. (2007). Project Appraisal Document on a Proposed Credit in the Amount of Sdr 57.8 Million (US $86.0 Million Equivalent) to the Goverment of Kenya for a Western Kenya Community Driven Development and Flood Mitigation Project.
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.

## Figures

FIGURE 1: CORRUPTION CONTROL VS AID

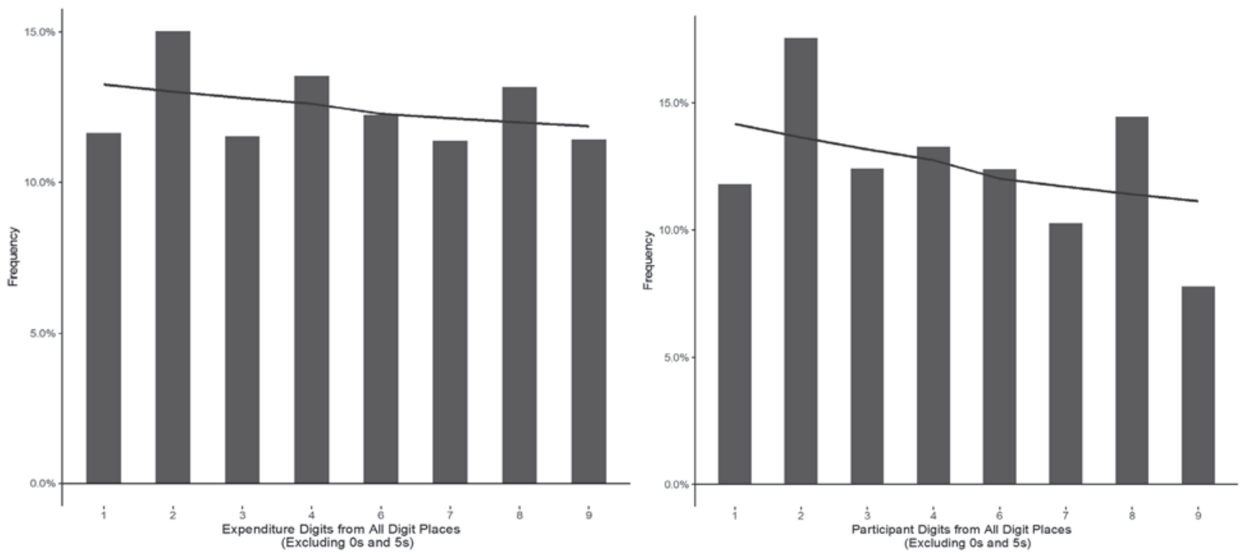



This figure plots the Worldwide Governance Indicator (WGI) control of corruption measure against log aid flows in 2019. WGI control of corruption measures "perceptions of the extent to which public power is exercised for private gain," standardized to mean 0 and standard deviation 1 (World Bank WGI, 2019); lower values correspond to lower controls and more corruption. Countries with worse corruption controls receive more aid. Of the $115 billion dollars of foreign aid to countries in these data, 92% of aid dollars flow to countries where corruption control is below the mean. The slope of the linear regression is -0.95, ($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.

FIGURE 2: ALL DIGIT PLACES BEYOND THE FIRST AGAINST BENFORD'S LAW FOR EXPENDITURE AND PATICIPANT DATA

A: Expenditure Data

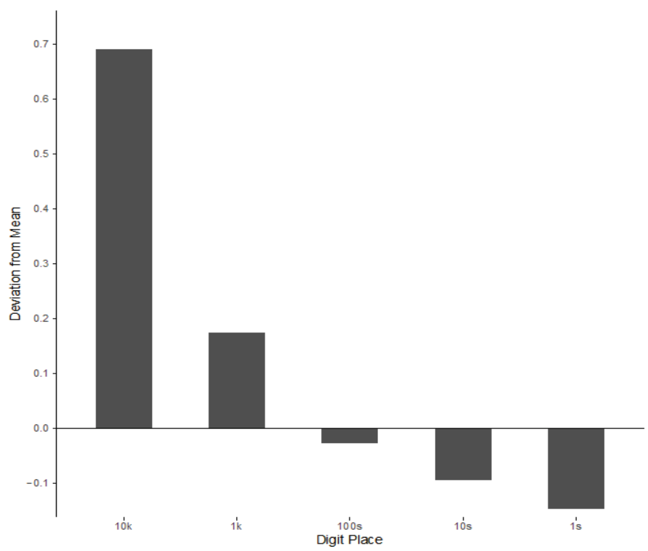
B: Participant Data



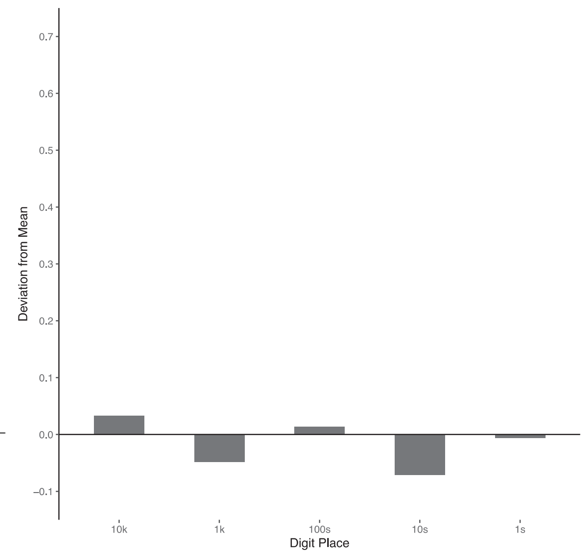

This figure presents all digits from beyond the first place, pooled, from expenditure data (Panel A) and participant data (Panel B) from all districts combined. The expected Benford's Law distribution is the solid line. Both tests are statistically significant, with $p = 3.9 \times 10^{-15}$; $n = 9371$ for the expenditure data (left) and $p = 5.7 \times 10^{-51}$; $n = 7385$ for the participant data (right). 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. Tests of rounding are presented in Appendix A.

FIGURE 3: PADDING TEST OF MONETARY INCENTIVES WITH MONTE CARLO SIMULATION

Panel A: World Bank Data



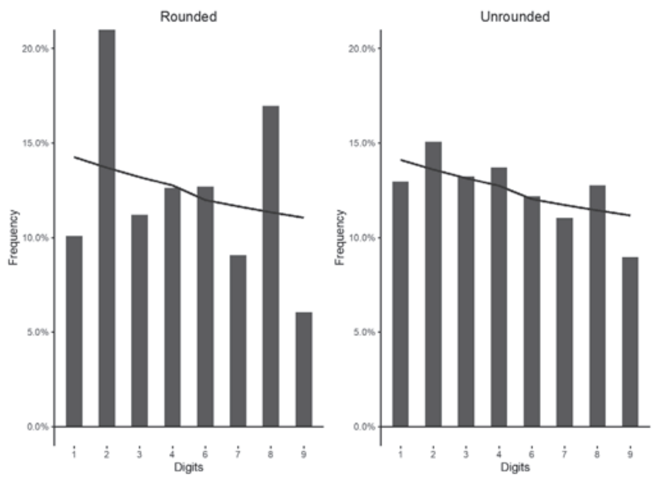

Panel B: Simulated Data



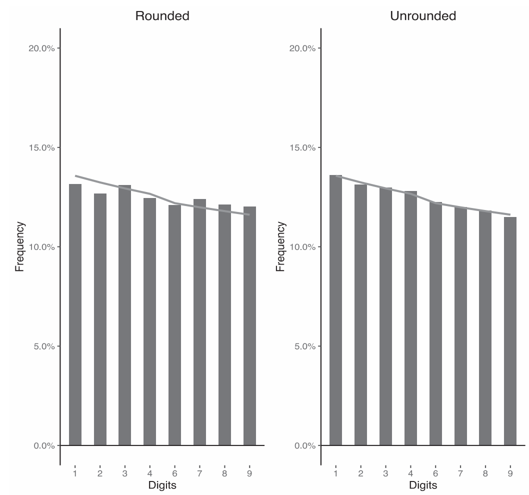

While traditional Benford's Law tests are performed from the left, i.e. first digit, second digit etc, here we test digits from the right, i.e. ones, tens, hundreds, which allows us to observe monetary incentives to pad more valuable digits. We compare the mean in each digit place from the right to the Benford expected mean in each sector. Zero reflects conformance to the Benford expected mean, and positive values indicate the mean 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. We perform a Monte Carlo simulation of Benford-conforming datasets and compare our observed statistics to the simulated statistics to produce p-values. Compared to a sample of 100,000 simulations, using data from all sectors, we observe the following statistics for the World Bank Data: 10,000s place ($p = 1.0 \times 10^{-5}$), 1,000s ($p = 2.3 \times 10^{-4}$), 100s ($p = 0.33$), 10s ($p = 0.10$), 1s ($p = 0.061$). Panel B shows simulated Benford-conforming data with 10,000 observations. No such pattern emerges.

FIGURE 4: UNPACKING ROUNDED AND UNROUNDED DIGITS IN PARTICIPANT DATA

PANEL A: WORLD BANK DATA

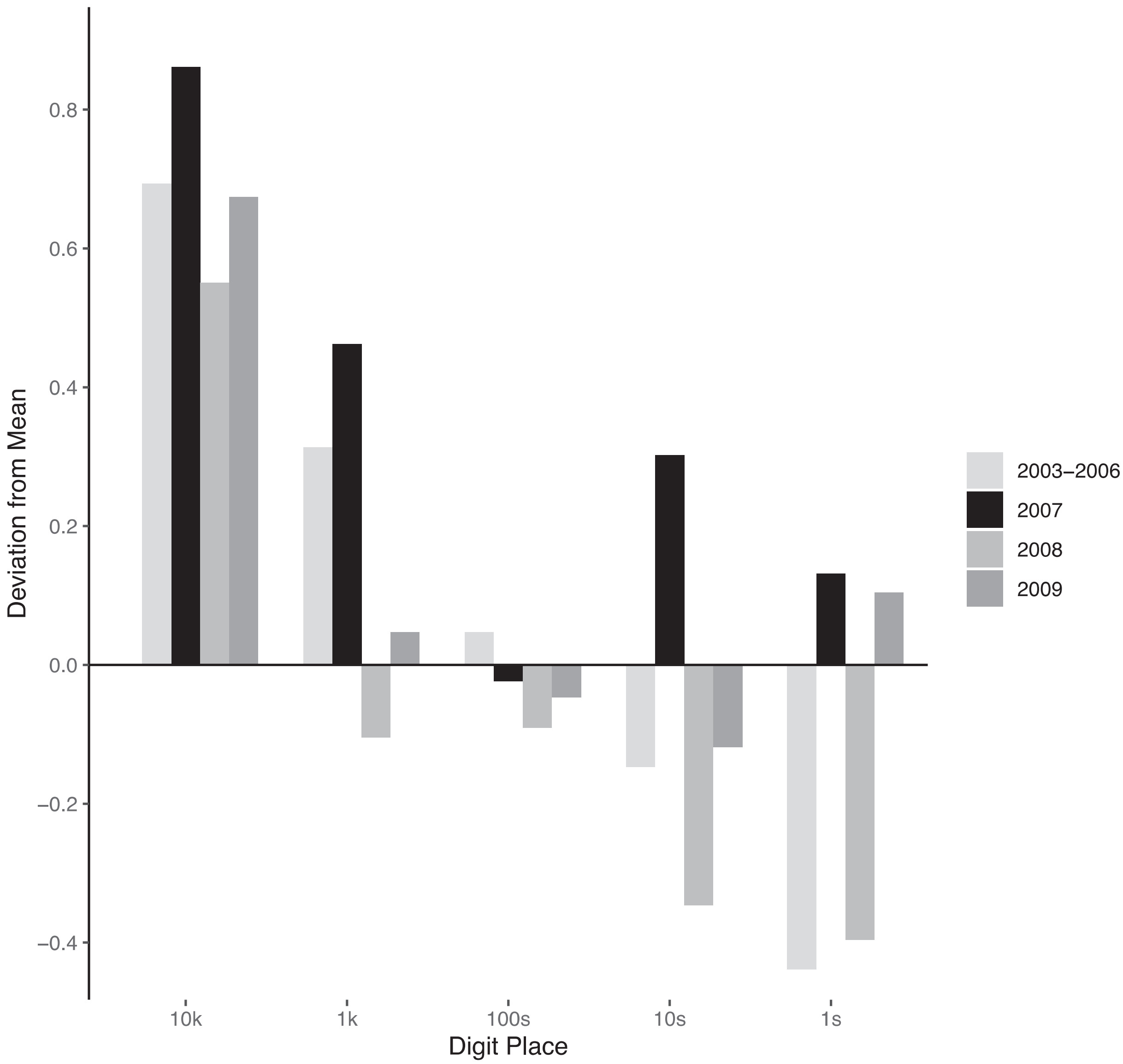



PANEL B: SIMULATION




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, $p = 1.9 \times 10^{-6}$; $n = 4410$. 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.

FIGURE 5: ELECTION YEAR EFFECTS IN EXPENDITURE DATA




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 thousands place, $p = 0.0001$; one thousands place, $p = 0.0001$).

## Tables

TABLE 1: EXPECTED DIGIT FREQUENCIES UNDER BENFORD'S LAW

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

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

TABLE 2: COMPARISON OF SINGLE DIGIT TESTS TO NEW ALL DIGITS TEST WITH SIMULATED DATA

| Panel A: Single Digit Tests | | | | | | | |
|---|---|---|---|---|---|---|---|
| | District A | District B | District C | District D | District E | District F | All Districts |
| First Digits | .0247 | .3849 | 0.2607 | 0.9681 | 0.5627 | 0.7321 | 0.5259 |
| Second Digits | **0.0009345** | 0.3319 | 0.3817 | 0.2 | 0.2157 | 0.1086 | 0.1378 |
| Third Digits | 0.02922 | 0.05461 | 0.4149 | 0.1289 | 0.06716 | 0.3711 | 0.1919 |
| Last Digits | **0.002284** | 0.08462 | **0.0002037** | 0.06975 | 0.0299 | **0.001027** | **1.778e-11** |
| *n* (per test) | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 6,000 |
| Panel B: All Digit Places | | | | | | | |
| | District A | District B | District C | District D | District E | District F | All Districts |
| All Digit Places | **6.885e-5** | **0.003257** | **.03098** | 0.1345 | 0.1993 | 0.411 | 0.2367 |
| *n* | 5,503 | 5,410 | 5,507 | 5,458 | 5,488 | 5,511 | 32,877 |

This table shows the result of simulated data, where single digits are tested separately (Panel A) and simultaneously (Panel B). Bolded values are statistically significant, corrected for multiple testing with a Bonferroni correction (0.05 divided by 4 tests in panel A for a significance level of 0.0125). Only districts A, B, C have manipulated data, but single-digit testing fails to detect this, while also inappropriately flagging District F in a last-digits test. Districts A, B, and C, which have manipulated data, are correctly identified by an all digit places test.

## TABLE 3. 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 | ■ | ■ | ■ | □ | ■ | ■ | □ | ■ | ■ | □ | ■ | ■ |
| 2B | All Digit Places Beyond the First: Participant | ■ | ■ | ■ | ■ | ■ | ■ | ■ | □ | □ | ■ | □ | ■ |
| 3 | Padding Valuable Digit Places | ■ | □ | ■ | □ | □ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| 4 | Unpacking Rounded Numbers: Participant | ■ | ■ | ■ | ■ | □ | ■ | ■ | □ | ■ | ■ | □ | ■ |
| 5 | Election Year Effects: Expenditure | □ | □ | ■ | □ | □ | ■ | ■ | □ | ■ | ■ | □ | ■ |
| A1 | Rounding Digits: Expenditure | ■ | ■ | □ | ■ | □ | □ | □ | ■ | □ | □ | □ | NA |
| A2 | Repeating Numbers: Expenditure | ■ | □ | □ | ■ | ■ | □ | □ | □ | □ | □ | □ | NA |
| A3 | Sector Effects: Expenditure | ■ | ■ | ■ | ■ | ■ | □ | ■ | ■ | □ | □ | □ | ■ |
| A4 | First Digit: Expenditure Data | ■ | ■ | □ | ■ | ■ | □ | ■ | □ | ■ | □ | ■ | □ |
| A5 | Digit Pairs: Participant | □ | ■ | ■ | □ | ■ | ■ | □ | ■ | □ | □ | □ | ■ |
| | Number of Significant Tests p < 0 .005 (Out of 10) | 8 | 7 | 7 | 6 | 6 | 6 | 6 | 5 | 5 | 4 | 3 | |

■ p < 0.005 □ p >= 0.005

We run 10 digit tests on each of 11 districts. These tests are chosen to analyze different, non-overlapping aspects of the data. Given the large number of tests, a Bonferroni correction is used to establish 0.005 as the acceptable *p* – value for our tests. Failed tests at the 0.005 level are indicated in dark grey. Two tests, which compare rounding and repeats *across* districts, are not applicable for all districts combined. We tabulate the number of significant tests for each district in the bottom row. Exact *p*-values for each test are presented in Appendix Table A1.

TABLE 4. DIGIT TESTS BY DISTRICT COMPARED TO WORLD BANK INT FORENSIC AUDIT RESULTS

| | Digit Tests (Number Failed 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 |

This table shows the number of digit analysis tests failed, 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 a forensic audit. 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 $p = 0.023$, $t$-statistic 4.33. The detailed results of each digit test are presented in Table 3 and Appendix A. The source for the INT forensic audit data is (World Bank Integrity Vice Presidency, 2011).