

# Auditing Public Budget Modifications in Mexico: An Isolation-Based Anomaly Detection Approach<sup>1</sup>

Gustavo Guajardo<sup>2</sup> and Francisco Cantú<sup>3</sup>

## Abstract

The misappropriation of public resources constitutes a major problem worldwide, with public officials often deviating from their budgets and redirecting funds towards questionable purposes. This letter adapts isolation-based anomaly detection (iForest), an unsupervised and nonparametric algorithm, to identify red flags in public budget modifications. We demonstrate the advantages of iForest by analyzing two million expenditure reports and audit data of the Mexican government (2010-2019). We find that audit results are positively associated with the anomaly scores produced by iForest, and iForest outperforms alternative approaches. Additionally, we show that anomalous cases exhibit higher rates of sanctioned officials and qualitative evidence of malfeasance. This approach could help auditors and scholars identify malfeasance risk when audit data is not available, reliable, or sufficient.

**Keywords:** Anomaly detection, corruption, audits, machine-learning, Mexico

---

<sup>1</sup> Please do not circulate without authors' consent.

<sup>2</sup> Ph.D. candidate, Rice University (gguajardo@rice.edu)

<sup>3</sup> Professor of Political Science, University of Houston (fcantu10@central.uh.edu)

The misappropriation of public resources constitutes a significant problem worldwide, with public officials often deviating from their budgets and redirecting funds towards political campaigns, salaries, and lavish spending (Fan et al. 2010; Fantaye 2004; Fisman and Golden 2017). Audits of public spending offer a way to uncover these practices, and social scientists have leveraged these data to study the causes and consequences of malfeasance (Arias et al. 2019; Avis et al. 2018; Berliner and Wehner 2022; Bobonis et al. 2016; Chong et al. 2015; Ferraz and Finan 2011; Larreguy et al. 2020; Olken 2007).

Despite their importance, public audits are rare and face at least three limitations. First, distinguishing between red flags and legitimate budget modifications is challenging since not all modifications are inherently malicious. Second, a comprehensive analysis of government spending is time-consuming and resource-intensive. Auditors often face tight deadlines with limited budgets and staff (Carslaw and Mason 2012; Gallego et al. 2021). Under these constraints, auditors must limit their analysis to a subset of cases using an arbitrary criterion. Third, audits have scope limitations and are not legally allowed to inspect certain cases.<sup>4</sup> Moreover, oversight institutions can be subjected to external pressures to politicize their work and experience efforts to weaken their independence and capacities (Vera and Pozsgai-Alvarez 2022).

This letter adapts isolation-based anomaly detection (iForest), an unsupervised and nonparametric machine-learning algorithm commonly used to detect fraudulent credit card transactions and breaches to network security (Liu et al. 2012; Bandaragoda et al. 2018), as a risk-

---

<sup>4</sup> For example, both Mexico's and Brazil's supreme audit institutions only revise federal funds. Moreover, large Brazilian cities are considered ineligible (Odilla and Rodriguez-Olivari 2022).

based approach for identifying red flags in public budget modifications.<sup>5</sup> This approach can process vast amounts of data while only requiring information on budget modifications within government agencies. To illustrate, we use data from public audits and expenditure reports of agencies within the Mexican government (2010-2019). We provide three types of supporting evidence: First, we find that anomaly scores produced by iForest are positively associated with irregularities found by auditors. Second, iForest outperforms alternative selection criteria and unsupervised algorithms in identifying irregularities. Third, we show that anomalous cases exhibit higher rates of sanctioned public officials, state-level victims of corruption, and qualitative evidence of malfeasance.

With this study, we build on existing research in political science that uses machine learning to identify corrupt activities (Ash et al. 2021; Fazekas and Kocsis 2017; Gallego et al. 2021; Montgomery et al. 2015) and provide a supplementary tool for auditing agencies with limited resources. Additionally, this approach could help scholars monitor and quantify malfeasance risk within the government where audit data is not available, reliable, or sufficient.

---

<sup>5</sup> Similar approaches have been used to detect fraud in the private sector (Kokina and Davenport 2017; Ucoglu 2020).

## 2. Identifying anomalies in budget modifications

The main challenge in identifying a potential diversion of public resources is distinguishing malfeasance from legitimate budget modifications. Legitimate budget modifications redirect resources from lower-priority areas to cover large-scale unanticipated shocks, such as inflation, international crises, or natural disasters. Due to the scale of these shocks, budget adjustments should follow a similar pattern across expenditure categories and government agencies. By contrast, we argue that malfeasances should be rarer, localized, and less likely to have a general pattern across agencies and expenditure categories. They are triggered not by major shocks but by unjustified and discretionary decision-making. For example, public officials can coordinate with contractors to artificially inflate costs in infrastructure projects (Kenny 2009). Public officials can also accumulate personal wealth by embezzling public funds (Fantaye 2004). In Mexico, for instance, resources from social programs are often used to fund political propaganda.<sup>6</sup> We expect the data-generating process of malfeasances to be distinct from that of legitimate budget modifications.

To trace malfeasance in budget modifications, we use the anomaly detection algorithm iForest (Liu et al. 2012). The primary focus of iForest is identifying anomalies, which we argue have similar properties to malfeasances. Anomalies have two properties. First, they constitute a

---

<sup>6</sup> See articles from *Expansión Política* (<https://politica.expansion.mx/mexico/2020/09/07/por-corrupcion-94-municipios-perdieron-el-90-de-gasto-social-informe>), IMCO (<https://imco.org.mx/presupuestos-locales-mala-planeacion-o-gastos-inesperados/>), and *Animal Político* (<https://www.animalpolitico.com/2017/02/desvios-millonarios-programas-cuenta>).

minority of observations, and second, they come from a unique data-generating process compared to the rest of the observations. In contrast to conventional outlier detection algorithms, which measure the distance between each observation and the mean of a cluster or the entire dataset, iForest estimates the susceptibility of each observation to be separated from the rest without any distributional assumptions (Song et al. 2007; Liu et al. 2012).<sup>7</sup>

The unsupervised algorithm follows three general steps.<sup>8</sup> First, iForest builds decision trees using random subsets of observations and variables.<sup>9</sup> A variable is randomly chosen, along with a threshold for that variable's range. Second, the threshold splits observations into two groups based on whether they have lower or higher values. The process is repeated recursively for each new group until the pre-specified number of trees is reached.<sup>10</sup> The third step ranks all data points according to the number of necessary splits required to isolate each observation and assigns them an *anomaly score*. Anomaly scores reflect how anomalous an observation is compared to the rest of the covariates in the dataset. Observations requiring fewer splits to be isolated from the rest are considered more anomalous. Conversely, observations that require more splits to achieve isolation are considered to be less anomalous. Figure 1 visually illustrates this intuition, where dots

---

<sup>7</sup> The assumptions are that all attributes in the dataset contribute equally to anomaly detection and there are no conditional anomalies (Song et al. 2007; Liu et al. 2012).

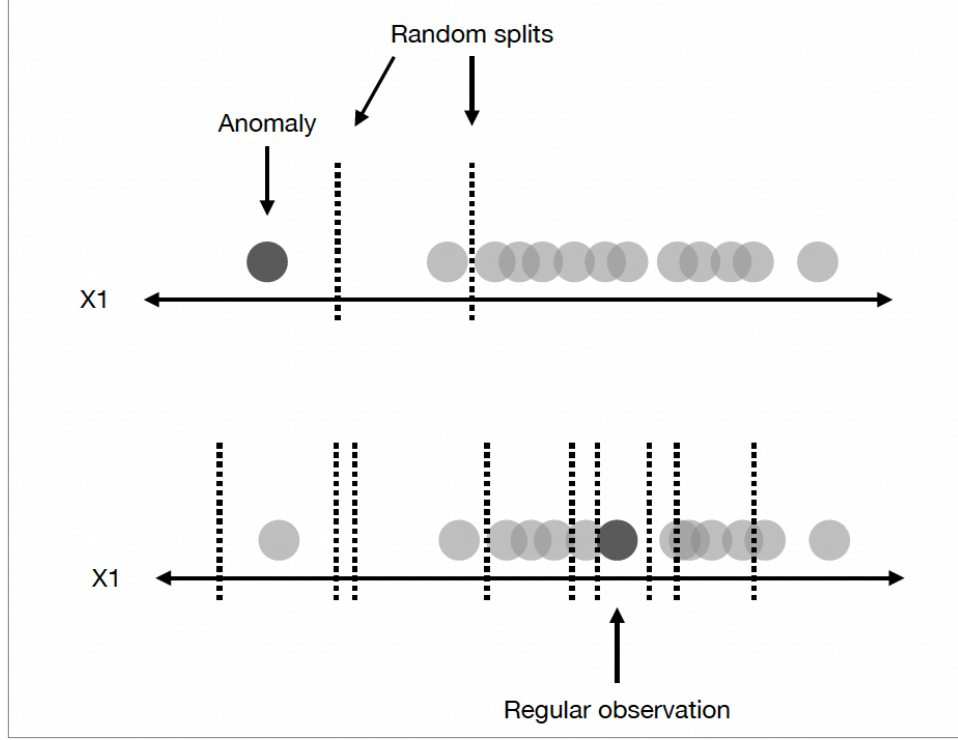
<sup>8</sup> Find more detailed step-by-step in the appendix section 1.

<sup>9</sup> Decision trees split the data based on simple true/false statements.

<sup>10</sup> iForest requires two key hyperparameters: the number of trees, and the size of the training data subsample. As the size of the subsample increases, isolation forest detects more reliably. Liu et al (2012) found that path lengths converge well before 100 trees.

represent observations in the feature space (black line) of variable  $X_1$ . Compared to a regular observation (lower panel), an anomalous observation (upper panel) will take fewer splits to achieve isolation.

Figure 1. Isolation forest visualized



Equation 1 shows the estimation of anomaly scores, where  $E(h(x))$  is the expected number of edges in the trees for a particular data point “x”,  $h(x)$  is the path length of “x” (or the average number of trees required for isolation), and  $c(n)$  is the normalization constant for a dataset of size “n”.

$$Anomaly\ score = 2^{-\frac{E(h(x))}{c(n)}} \quad (1)$$

### 3. Data and methods

We apply isolation-based anomaly detection to a fine-grained dataset of Mexican government expenditure reports. These expenditure reports are freely accessible to the public in the government's online open budget portal and contain information on over 2.6 million budget modifications across 1,379 subnational government ministries and over 12,000 agencies between 2010 and 2019.<sup>11</sup> Reports specify the yearly initial and modified budget by expenditure category of a government agency within a ministry.

The process of modifying the budget in Mexico works as follows. Every year, the Secretary of Finance (SHCP) publishes last year's expenditures. A new budget draft is created as a joint effort among ministries, agencies, and the SHCP. Congress eventually approves the proposed budget. Throughout the year, agencies can request budget modifications, with most cases handled within their administrative office.<sup>12</sup> The criteria for approving these requests are vague and highly discretionary. In some cases, modifications are approved as long as they follow the National Development Plan's goals, which contain generic goals such as "*helping Mexico achieve its fullest potential*" (Núñez 2017). Budget changes in Mexico are commonplace. Figure 2 shows the programmed and observed expenses of the Mexican government (2010-2020), as well as the

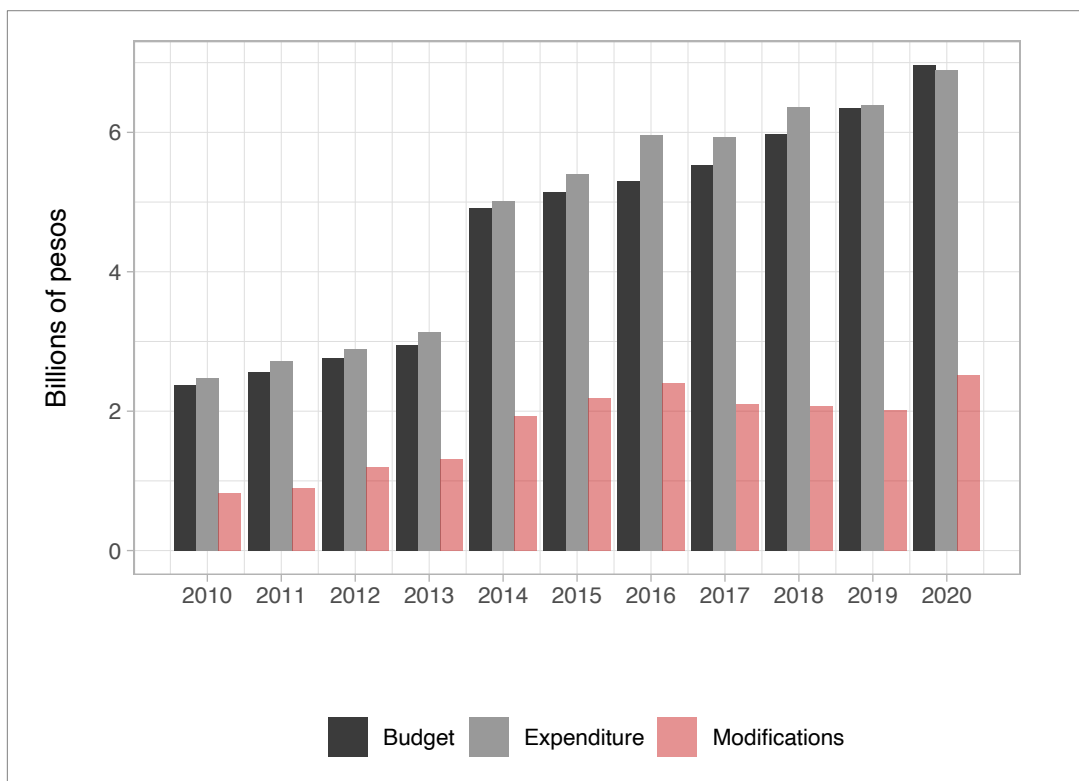
---

<sup>11</sup> Transparencia Presupuestaria (<https://www.transparenciapresupuestaria.gob.mx/Datos-Abiertos>).

<sup>12</sup> The Secretary of Finance keeps track of real-time expenditure. Budget modifications are made through a procedure called "*ajustes presupuestarios*," and requests are submitted to each agency's administrative office. The Secretary of Finance handles the approval of larger budget modifications.

money spent on a category different from the initial one—a budget modification (in red). Budget modifications can represent up to 30 percent of the total budget in a fiscal year.

Figure 2. Budget modifications of the Mexican government



*Note:* Budget modifications (red) comprise the absolute value of the sum of all budget modifications, both positive (overspending) and negative (underspending).

To validate our approach, we leverage a dataset containing 25,000 audits within 600 government ministry-years (2010-2019) conducted by Mexico’s supreme audit institution (ASF),<sup>13</sup>

<sup>13</sup> Publicly available in the online portal of the ASF (<http://www.asfdatos.gob.mx/>).



data on sanctions to public officials,<sup>14</sup> and data on state-level perceptions of corruption from a nationally representative survey.<sup>15</sup> Our primary measure of malfeasance is the reported irregularities uncovered by auditors. Auditors quantify the amount of mishandled funds and refer to them as “irregularities”, commonly used as a proxy for malfeasance (Chong et al. 2015; Larreguy et al. 2020). The data from the ASF includes multiple audits to ministries at different levels. Figure A3 in the appendix shows that one in every five audits finds irregularities. To match these findings to our database on public expenditures, we aggregated the agency-year level data to ministry-year level.

We use iForest to single out anomalies in budget modifications. We consider budget modifications (in Mexican pesos) to be the difference between the initial budget and the end of the year expenditure for each ministry ( $i$ ) and expense item ( $j$ ) at year ( $t$ ). We add 1 unit to the denominator to avoid dividing by zero.

$$Modification_{ijt} = \frac{(Spent_{ijt} - Assigned_{ijt})}{(Assigned_{ijt} + 1)} \quad (3)$$

Setting up the data for analysis requires three steps. First, choose a unit of analysis (agency, ministry, or state) and aggregate budget modifications to that respective level. For example, if the unit of interest is government ministries, we would aggregate budget modifications that initially

---

<sup>14</sup> Publicly available in the government’s transparency portal

(<https://www.plataformadetransparencia.org.mx/>).

<sup>15</sup> Survey of government quality and impact(ENCIG), publicly available online

(<https://www.inegi.org.mx/programas/encig/>).

came at the agency-year level or lower to ministry-year level. Second, identifying all unique categories of expenditure.<sup>16</sup> Third, restructuring the data so that each row represents a unit of analysis “i” in time “t”, and each column contains the budget modifications ( $Modification_{ijt}$ ) for all possible categories of expenditure (j). Under this setup, units are compared to one another, and information on budget modifications to specific expenditure categories is used to single out anomalies in the data.<sup>17</sup> We run iForest using the solitude package in R Studio (version 1.1.3). hyperparameter values were 150 trees and 256 subsamples.<sup>18</sup>

#### 4. Validity tests

In this section, we demonstrate the advantages of using iForest as a criterion for risk-based auditing. We proceed in three steps. First, we examine the association between the anomaly scores produced by iForest and irregularities found by auditors. Table 1 presents the result of regressing irregularities on anomaly scores. The unit of analysis is a government ministry-year, and we consider cases where we observe audit results. The first model considers the total amount of irregularities found in audits within the ministry-year. The second model uses a binary indicator for whether the observation was in the top anomalous decile (1) or not (0) to predict irregularities.

---

<sup>16</sup> To reduce the dimensionality of the data, we combine expenditure categories that share a theme and go well together. For example, we combined categories such as “Clothing and uniforms” and “Clothing, whites, protective clothing and sporting goods” into one.

<sup>17</sup> Find a visual example of the data in section 2 of the appendix.

<sup>18</sup> Appendix Figure A5 reports the results of hyperparameter sensitivity analysis, finding that estimates become consistent after the values chosen for the number of trees and sub sample size.

In columns 3-4, the outcome is coded as “1” if audits to the ministry found any irregularities and “0” otherwise. Results show a strong and statistically significant relationship across models, indicating that higher anomaly scores predict larger irregularities found by auditors.

One potential concern is that ministries with fewer resources present irregularities due to low administrative capacity. Table A1 in the appendix addresses those concerns by using the ministry’s budget—a proxy of administrative capacity—to predict the proportion of audits to the ministry that found irregularities, finding no significant relationship between them. Additionally, we run models with simulated anomaly scores as a placebo test to rule out a spurious relationship. Similarly, the model does not exhibit a significant relationship (Appendix Table A2).

Table 1. Irregularities uncovered by auditors and anomaly scores

	Irregularities		Irregularities found [0,1]	
	(1)	(2)	(3)	(4)
Anomaly scores	0.65*** (0.11)		2.55*** (0.79)	
Top anomalous decile		0.16*** (0.04)		0.50** (0.21)
Constant	0.04 (0.03)	0.07*** (0.03)	1.02*** (0.13)	1.15*** (0.12)
Observations	600	600	600	600
R <sup>2</sup>	0.06	0.03		
Log Likelihood			-300.25	-304.24

*Note:* OLS (1-2) and logistic (3-4) regressions with standard errors in parenthesis. Outcomes are the total amount of irregularities found in audits to the ministry (1-2) and a binary measure for irregularities found/no irregularities found (3-4). \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Our second step compares the performance of iForest against alternative approaches. By construct, the available data overestimates the number of irregularities. The process of aggregating audit results to ministry-level, which provides the cleanest match between the audit data and the public expenditures data, leads to a dataset where 80 percent of ministries present irregularities in at least one of their audits. To avoid inflating the detection accuracy of iForest, we use synthetic data (simulation details in the Appendix section 8). We generate synthetic cases resembling observations without irregularities by using the moments of the original distributions and illustrate this process in Figure 4.<sup>19</sup> This generates three different subsets of data: units with irregularities (gray), units without irregularities (white), and the synthetic data resembling units without irregularities (dotted lines). We merge the last two categories into one (columns 3-4) and randomly draw observations from the two resulting groups to create four sets of 1000 samples, each with a different rate of observations with irregularities—2, 5, 10, and 20 percent.

The premise of this exercise is that auditors face constraints and can inspect 10 percent of cases. We compare the detection accuracy of five approaches: random audits, naïve audits, k-means clusters, local outlier factor (LOF), and iForest. Random audits make an aleatory selection of 10 percent of the observations. Naïve audits select 10 percent of the observations with the largest budget modifications.<sup>20</sup> The rest are unsupervised algorithms. K-means groups observations into

---

<sup>19</sup> We draw from the gamma distribution with the mean and SD of the original variables. We estimate the shape parameter ( $\alpha$ ) through  $\alpha = \left(\frac{\mu}{\sigma}\right)^2$  and the scale parameter ( $\theta$ ) through  $= \frac{1}{(\mu)/(\sigma)^2}$

<sup>20</sup> We calculate this value through the summation of the absolute values of all budget modifications for each observation ( $\sum_{i=1}^n |\rho_i|$ ).

clusters based on their features and considers observations far from clusters as outliers (Ding et al. 2009).<sup>21</sup> We select the 10 percent with the largest distance from the center of their cluster. LOF detects outliers by comparing the local density of a data point to that of its neighbors (Breunig et al. 2000). For LOF and iForest, we select the sample observations with the highest anomaly scores. While supervised approaches—such as gradient-boosted classifiers—have been shown to have high prediction accuracy (Ash et al. 2021), they require high-quality data to train and are prone to overfitting. We focus on unsupervised algorithms since our goal is overcoming the limitations of audit data, which is rare, has varied quality, and is often constrained to inspect only some cases in certain times.

Figure 5 compares the proportions of correctly identified irregularities (or true positive rates) of each approach for our four sets of 1000 samples. Each set has a different percentage of irregular observations. We find that iForest is better at identifying irregularities than alternative approaches. The overperformance of iForest becomes more evident as the samples contain more irregularities. For instance, in samples with 20 percent of observations with irregularities, iForest, on average, correctly identifies them 60 percent of the time. On the other hand, naïve audits, the second most effective approach in our comparison, correctly identify irregularities around 45 percent of the time.

---

<sup>21</sup> We compute the number of clusters using the Elbow method with the RclusTool package (version 0.91.5).

Figure 4. Constructing samples with different rates of irregularities

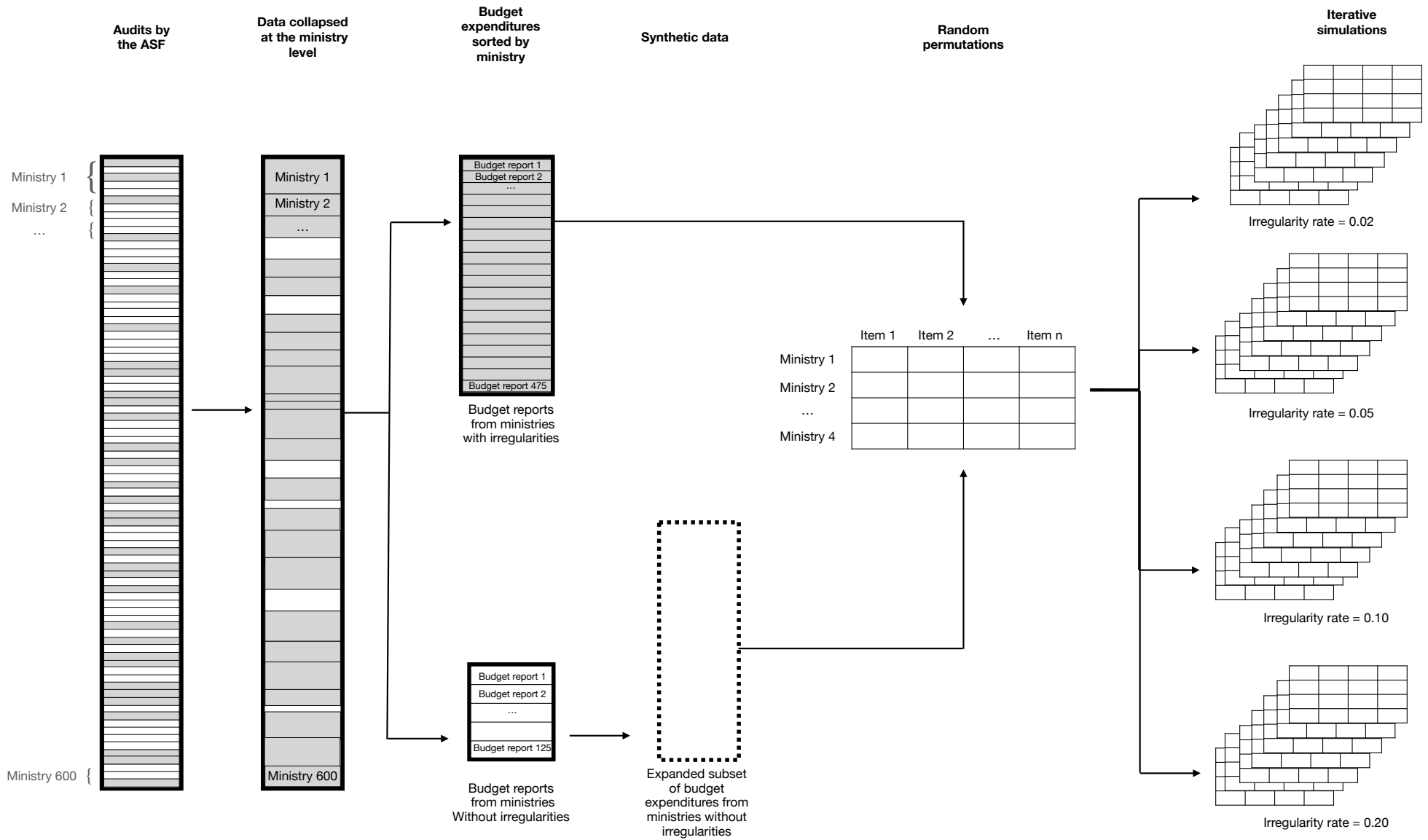
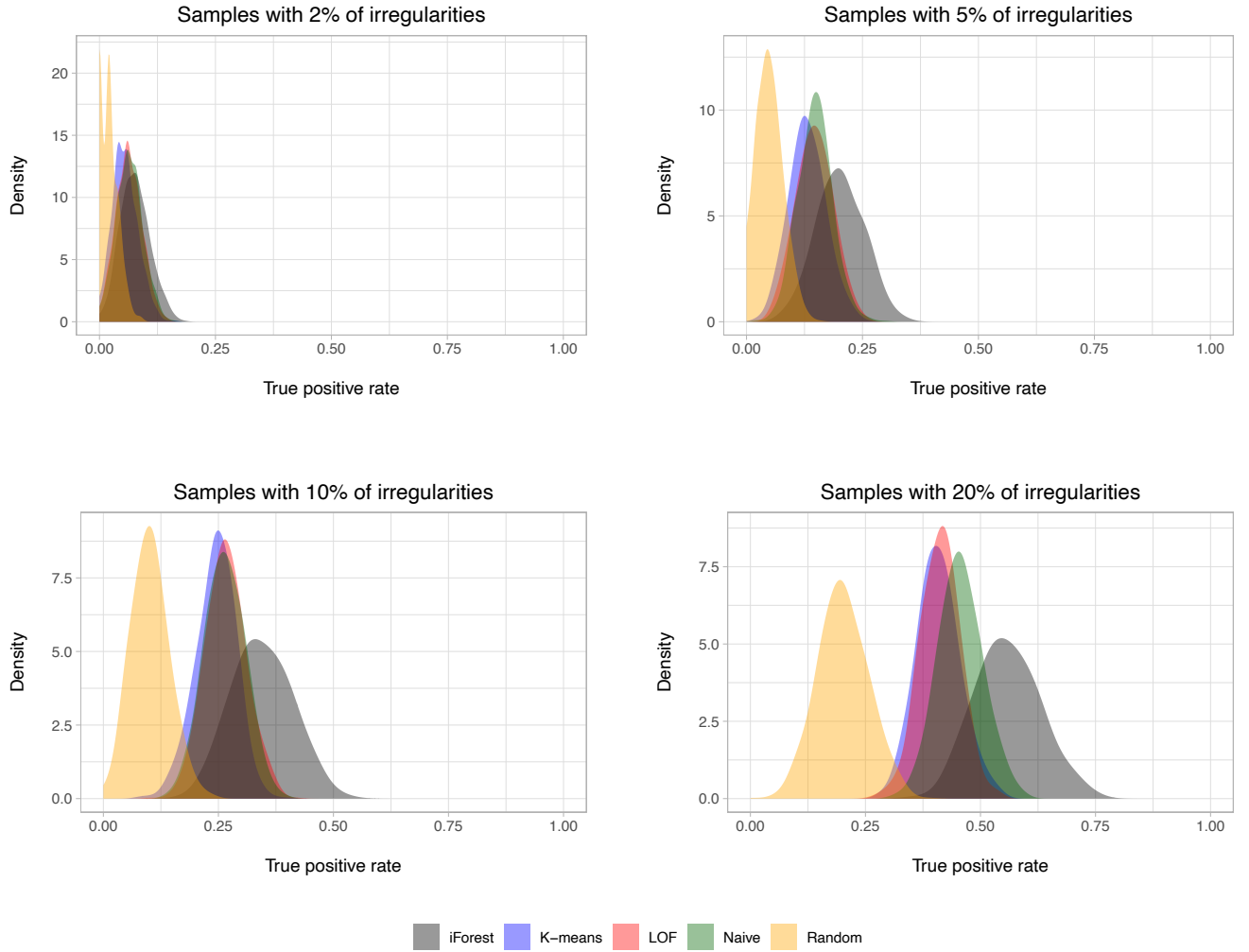


Figure 5. Simulation results



*Note:* Rates of correctly identified irregularities (true positive rate) for the 1000 simulated samples under each approach.

Next, we apply iForest to our fine-grained data on budget modifications in Mexico. We use three types of supporting evidence: data on sanctioned public officials by ministry, state-level victims of corruption, and a collection of malfeasance scandals articles (see these news reports in appendix Table A3). Figure 6 plots the predicted anomaly scores of government ministries against their number of sanctioned public officials, showing a positive correlation between them ( $corr =$

0.4).<sup>22</sup> The Institute for Social Security and Services for State Workers (ISSSTE) presents the highest anomaly score and number of sanctioned officials. Together with the Mexican Social Security Institute (IMSS), social security institutions are frequently linked to irregular activity, duplicated payments, embezzlement, and inflated prices for pensions, medicines, and equipment.<sup>23</sup> Another case with a high anomaly score is Pemex, Mexico's public oil company, which consistently faces allegations of irregular purchases, bribery, and embezzlement.<sup>24</sup> These ministries also have the highest numbers of corruption complaints submitted to the Ministry of Civil Service.<sup>25</sup> The anomaly scores are also consistent with evidence that the legislative branch diverted over 30.6 million dollars to shell companies between 2013 and 2018.<sup>26</sup> On the other hand, agencies like the Ministry of Culture, which have very low anomaly scores, present low levels of accusations and sanctioned officials. We find cases that deviate from the general pattern. For example, the Ministry of Tourism (SECTUR) presents low anomaly scores despite having high

---

<sup>22</sup> Figure A7 in the appendix ranks ministries based on their predicted anomaly score. The number of sanctioned officials ranged from 0 to 238.

<sup>23</sup> The General Prosecutor accused ISSSTE officials of fraud for 831 million and investigations found that between 2014 and 2019, both ISSSTE and IMSS embezzled over 4 billion pesos (see Table A3 in appendix).

<sup>24</sup> Four former heads were prosecuted in 2021, and Pemex has received 1.1 million US dollars in sanctions over the last 20 years (see Table A3 in appendix).

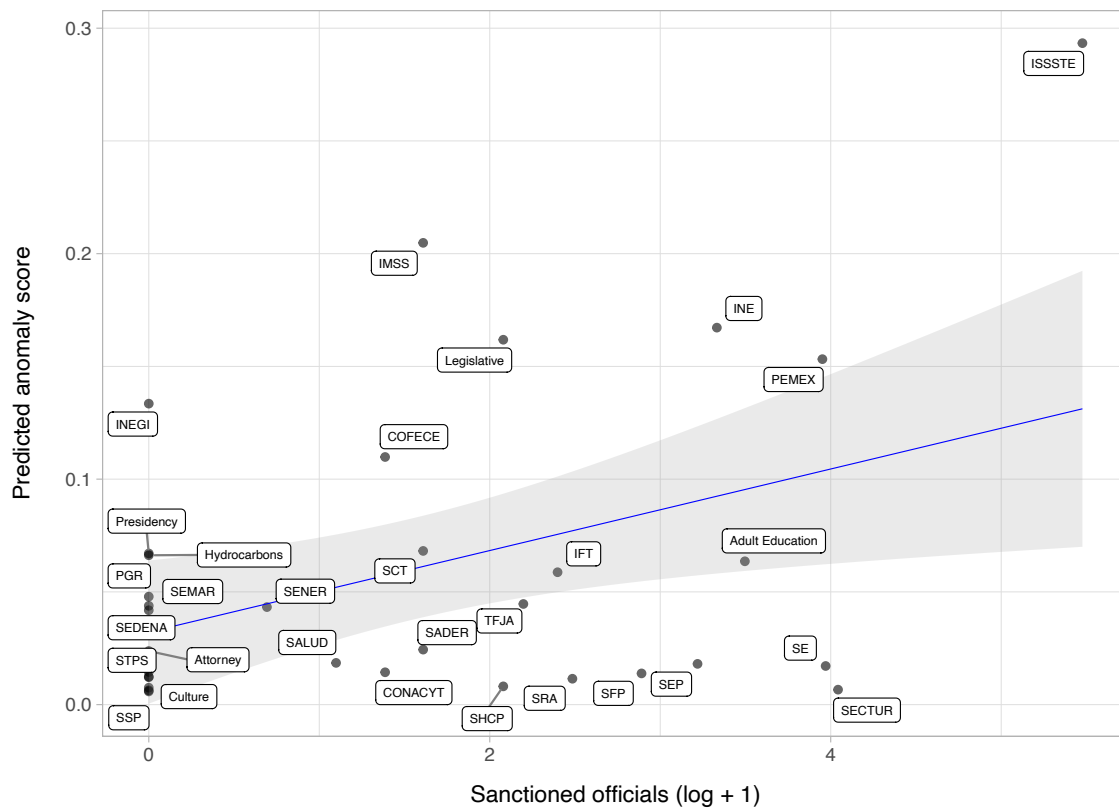
<sup>25</sup> Between 2018 and 2022, the SFP listed IMSS, ISSSTE, and Pemex as the ministries with the most complaints (see Table A3 in the appendix).

<sup>26</sup> See Table A3 in the appendix.



rates of sanctioned officials. The inconsistency could be an artifact of the sanctions measure including other types of wrongdoing, such as administrative negligence. Figure A8 and Table A4 in the appendix also showcase a positive and significant relationship between anomaly scores and state-level victims of corruption.<sup>27</sup>

Figure 6. Sanctioned officials and predicted anomaly scores for Mexican ministries



*Note:* Y-axis shows the predicted anomaly score, obtained through a beta regression predicting anomaly scores with ministry dummy variables. X-axis is the number of sanctioned public officials (log + 1) in said ministry as of 2022.

<sup>27</sup> The number of respondents in the survey of government quality and impact (ENCIG) that experienced corruption while interacting with a public official.

To provide further insight, Table A5 in the appendix lists the agencies with the highest anomaly scores and their most anomalous expenditure categories. Social security (ISSSTE) in Mexico City stands out, with medicines, medical equipment, and pensions receiving the highest anomaly scores for the agency. The years when ISSSTE was found to be most anomalous (2014-2019) were precisely those where investigations found evidence of fraud in the acquisition of medical equipment, inflated pensions, and embezzling of billions of pesos.<sup>28</sup> The federal police, accused of embezzling over 2.657 million pesos through simulated contracts and inflated prices in public-security equipment and vehicles, presents high anomaly scores on public-security expenditures for 2013, the year that investigators mark as the start of the illicit behavior.<sup>29</sup> The *Instituto Politécnico Nacional* (IPN), which presented high anomaly scores for third-party services, has recently been found to have created fake positions and third-party contracts in order to embezzle funds for over a decade.<sup>30</sup> The decentralized organ of Prevention and Social Rehabilitation was also found to have had its funds sacked by embezzlement schemes put in place by García Luna (former Secretary of Public Safety) and his officials.<sup>31</sup>

---

<sup>28</sup> See Table A3 in the appendix.

<sup>29</sup> See Table A3 in the appendix.

<sup>30</sup> Ibidem.

<sup>31</sup> Ibidem.

## 5. Conclusion

Machine-learning techniques can assist auditors and provide scholars with red flags in public budget modifications. This letter proposed isolation-based anomaly detection (iForest) as a selection standard for risk-based auditing. This approach has the appeal of only requiring information on budget modifications within government agencies. We used data from public audits and expenditure reports of agencies within the Mexican government, finding that anomaly scores produced by iForest are positively associated with auditors' findings. We also found that iForest outperformed alternative approaches, and qualitative evidence suggests that instances with recurrent corruption scandals exhibit higher anomaly scores. This approach can serve as an alternative tool for social scientists studying malfeasance in government.

Our findings also present opportunities for future research. We stress that this methodology should be considered a tool for assessing malfeasance risk in the absence of sufficient or reliable audit data and a complement to an auditors' expertise, not a substitute for audits. This approach has implications for research on the effectiveness of accountability mechanisms and policy. In line with similar studies (Ash et al. 2021), our findings underscore the effectiveness of risk-based approaches over traditional randomized audits. Finally, this approach has the potential to provide insights into the dynamics of malfeasance within the black box of government activity and the strategies to combat it.

## References

- Arias, Eric; Pablo Balan, Horacio Larreguy, John Marshall, and Pablo Querubín. 2019. “Information Provision, Voter Coordination, and Electoral Accountability: Evidence from Mexican Social Networks.” *America Political Science Review* 113, 475-498.
- Ash, Elliott, Sergio Galletta, Tommaso Giommoni. 2021. “A Machine Learning Approach to Analyze and Support Anti-Corruption Policy.” CESifo Working Paper, No. 9015.
- Avis, Eric; Claudio Ferraz, and Federico Finan. 2018. “Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians.” *Journal of Political Economy* 126, 1912-1964.
- Bandaragoda, Tharindu; Kai Ming Ting, David Albrecht, Fei Tony Liu, Ye Zhu, and Jonathan Wells. 2018. “Isolation-based anomaly detection using nearest-neighbors ensembles.” *Computational Intelligence* 34, 968–998.
- Berliner, Daniel and Joachim Wehner. 2022. “Audits for Accountability: Evidence from Municipal By-Elections in South Africa.” *Journal of Politics* 84, 1581-1594.
- Bobonis, Gustavo, Luis Cámara Fuertes, and Rainer Schwabe. 2016. “Monitoring Corruptible Politicians.” *American Economic Review* 106, 2371-2405.
- Breunig, Markus, Hans-Peter Kreigel, Raymond Tak Yan, and Jörg Sander. 2000. “LOF: Identifying density-based local outliers.” *ACM SIGMOD Record* 29, 2, 93–104.
- Carslaw, Charles; Pippin Sonja, and Richard Mason. 2012. “Are public sector auditors more effective than private sector audit firms when auditing governmental entities? Some evidence from United States.” *Public and Municipal Finance* 1, 49-57.
- Chong, Alberto, Ana De la O, Dean Karlan, and Leonard Wantchekon. 2015. “Does Corruption Information Inspire the Fight or Quash the Hope?” *Journal of Politics* 77, 55-71.

- Ding, Kexing, Xuan Peng, and Yunsen Wang. 2019. "A Machine Learning-Based Peer Selection Method with Financial Ratios." *Accounting Horizons* 33, 75-87.
- Fan, Simon; Chen Lin, and Daniel Treisman. 2010. "Embezzlement versus Bribery." NBER Working Paper 16542. <http://www.nber.org/papers/w16542>
- Fantaye, Dawit. 2004. "Fighting corruption and embezzlement in third world countries." *The Journal of Criminal Law* 68, 170-176.
- Fazekas, Mihaly, and Gabor Kocsis. 2017. "Uncovering High-Level Corruption: Cross-National Objective Corruption Risk Indicator Using Public Procurement Data." *British Journal of Political Science* 50, 155-164.
- Ferraz, Claudio, and Federico Finan. 2011. "Electoral Accountability and Corruption: Evidence from the Audits of Local Governments." *American Economic Review* 101, 1274-1311.
- Fisman, Ray, and Miriam Golden. 2017. *Corruption: What Everyone Needs to Know*. Oxford University Press.
- Gallego, Jorge, Gonzalo Rivero, and Juan Martinez. 2021. "Preventing rather than punishing: An early warning model of malfeasance in public procurement." *International Journal of Forecasting* 37, 360-377.
- Kenny, Charles. 2009. "Measuring Corruption In Infrastructure: Evidence from Transition and Developing Countries." *The Journal of Development Studies* 45, 314-332.
- Kokina, Julia, and Thomas Davenport. 2017. "The Emergence of Artificial Intelligence: How Automation is Changing Auditing." *Journal of Emerging Technologies in Accounting*, 14, 115-122.

- Larreguy, Horacio; John Marshall, and James Snyder. 2020. "Publicizing Malfeasance: When Local Media Structure Facilitates Electoral Accountability in Mexico." *The Economic Journal* 130, 2291-2327.
- Liu, Fei Tony; Kai Ming Ting, and Zhi-Hua Zhou. 2012. "Isolation-Based Anomaly Detection." *Transactions on Knowledge Discovery from Data* 6, Article 3, 1-39.
- Montgomery, Jacob; Santiago Olivella, Joshua Potter, and Brian Crisp. 2015. "An Informed Forensics Approach to Detecting Vote Irregularities." *Political Analysis* 23, 488–505.
- Núñez, Leonardo. 2017. *¿Y dónde quedó la bolita? Presupuesto de Egresos ficticio*. Editorial Aguilar.
- Odilla, Fernanda and Denisse Rodriguez-Olivari. Corruption control under fire: A brief history of Brazil's office for the comptroller general. In: *The Politics of Anti-Corruption Agencies in Latin America* (pp. 108-133). Routledge.
- Olken, Benjamin. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115, 200-249.
- Song, Xiuyao, Mingxi Wu, Christopher Jermaine, and Sanjay Ranka. 2007. "Conditional anomaly detection." *IEEE Trans. Knowl. Data Eng* 19, 631–645.
- Ucoglu, Derya. 2020. "Current Machine Learning Applications in Accounting and Auditing." *PressAcademia Procedia* 12, 1-17.
- Vera, Sofia and Joseph Pozsgai-Alvarez. 2022. The politics of building anti-corruption institutions: The case of anti-corruption agencies in Peru. Inside: *The Politics of Anti-Corruption Agencies in Latin America* (pp. 71-91). Routledge.

## Supplementary information

### Table of Contents

1. ISOLATION FOREST (STEP-BY-STEP).....	24
2. EXAMPLE OF DATA STRUCTURE.....	25
3. AUDITS WITHIN GOVERNMENT MINISTRIES.....	26
4. DISTRIBUTION OF ANOMALY SCORES.....	27
5. SENSITIVITY OF HYPERPARAMETERS .....	28
6. BUDGET SIZE AND IRREGULARITIES .....	29
7. PLACEBO TEST WITH SIMULATED ANOMALY SCORES .....	30
8. DETAILS ON THE SIMULATIONS.....	31
9. RANKING OF ANOMALY SCORES BY GOVERNMENT MINISTRY .....	32
10. COLLECTION OF NEWS ARTICLES ON MALFEASANCE SCANDALS .....	33
11. STATE-LEVEL PERCEPTIONS OF CORRUPTION.....	45
12. TOP ANOMALOUS AGENCIES AND CATEGORIES OF EXPENDITURE .....	47

## 1. Isolation forest (step-by-step)

iForest proceeds in 6 steps:

1. Randomly select instances (subsets) from the dataset based on the pre-specified number of trees and the training data subsample size. Subsamples are selected randomly without replacement.
2. Build decision trees using these samples as starting points.
3. For each internal node in the decision tree, randomly select a feature from the dataset and determine a threshold value for that feature.
4. Split observations into groups based on whether they are above or below the threshold value.
5. Repeat the process for each two groups, building decision trees for each group until all instances are isolated or a stopping condition is reached. The result is a set of decision trees, each of which represent a partition of the feature space in the data.
6. To detect anomalies the algorithm calculates the average depth of the instance in the decision trees. These are referred to as “path lengths”. The path length of a point “x” is basically the number of trees from the root to the end of the node. Observations that are isolated from the rest of the data will have deeper average depth and will be considered anomalies. Path lengths are normalized to create anomaly scores.



## 2. Example of data structure

Figures A1-A2 show visual examples of the data. Figure A1 shows a stylized example, where rows are ministry-years, and columns present the standardized budget modification (which could be negative, positive, or a zero) for a certain category of expenditure. These categories vary from salaries and bonuses to federal transfers, office supplies, machinery and vehicles, money for travel, construction expenses, etcetera. Figure A2 shows a screenshot of the actual data. Columns X1-X4 have the values of the standardized budget modifications for specific categories of expenditure, which are over 400 unique categories.

Figure A1. Data structure (stylized example)

ID	salaries	office supplies	machinery	...
Ministry X-Year 1	0.198	-0.032	0	
Ministry X-Year 2	0	-0.812	9.087	
Ministry X-Year 3	-0.004	0	0	
Ministry Y-Year 1	0	0.003	0	
Ministry Y-Year 2	2.334	0	0	
Ministry Y-Year 3	0	0	0	
...				

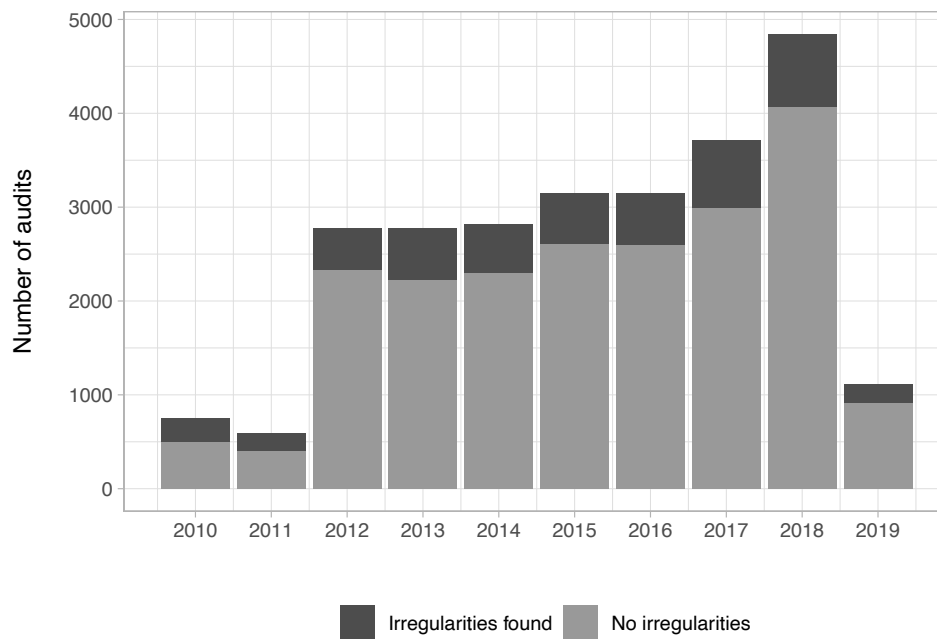
Figure A2. Data structure (actual data)

state	agency	year	anomaly_score	X1	X2	X3	X4
CDMX	AUDITORA SUPERIOR DE LA FEDERACION	2019	0.180034178	0	0.0000000000	0.000000	0.000000
CDMX	H. CMARA DE DIPUTADOS	2019	0.112579510	0	0.0006381375	0.000000	0.000000
CDMX	H. CMARA DE SENADORES	2019	0.120931692	0	0.0000000000	0.000000	0.000000
CDMX	AUTORIDAD EDUCATIVA FEDERAL EN LA CIUDAD DE ...	2019	0.339492440	0	-1.9226286000	-0.999624	-1.904628
CDMX	OFICIALA MAYOR	2019	0.004020823	0	0.0000000000	0.000000	0.000000
Aguascalientes	DELEGACION ESTATAL EN AGUASCALIENTES	2019	0.012078810	0	0.0000000000	0.000000	0.000000

### 3. Audits within government ministries

Figure A3 plots the count of audits within government ministries that found irregularities and those that did not find irregularities. One in every five audits found a positive amount of irregularities.

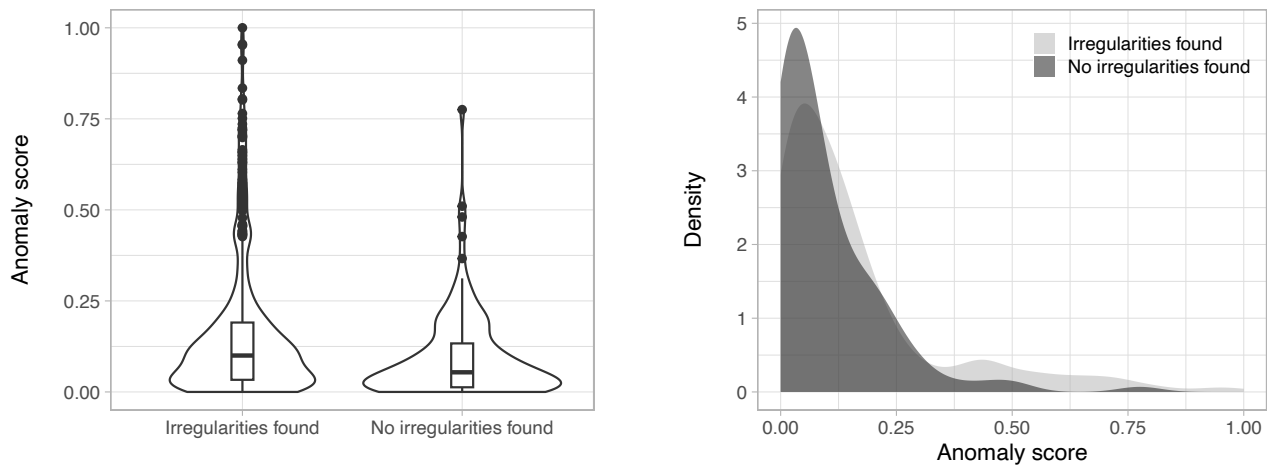
Figure A3. Audits within government ministries



## 4. Distribution of anomaly scores

Figure A4 plots the distribution of anomaly scores, which once normalized range between 0 and 1. These plots echo results from the regression table (Table 1) in the main text. Overall, anomaly scores are higher in ministries with dirty audit results, compared to those that found no irregularities.

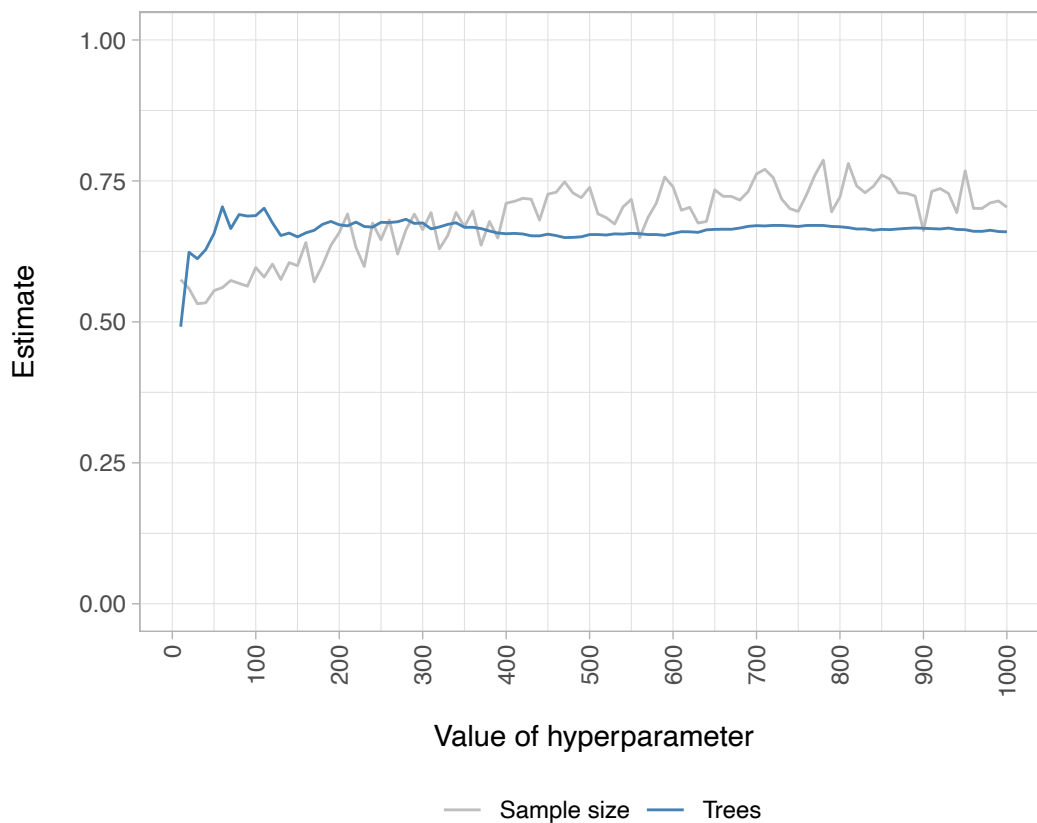
Figure A4. Distribution of anomaly scores (irregularities found vs no irregularities found)



## 5. Sensitivity of hyperparameters

Figure A5 plots the estimates of regressing irregularities found by auditors on anomaly scores across different values of the two hyperparameters: the number of trees and the size of the training data subsample. Results show that findings are remarkably consistent even low values of each hyperparameter—always positive, and between the values of 0.53 and 0.78. Estimates become more stable after trees = 150 and sample size = 250.

Figure A5. Hyperparameter sensitivity



## 6. Budget size and irregularities

A potential concern is that ministries with fewer resources could exhibit more irregularities due to low administrative capacity, and not wrongdoing. Table A1 addresses those concerns by using the size of the budget (log) to predict the proportion of audits to the ministry that found irregularities. We use budget size (the log of the total budget in pesos) as a proxy of administrative capacity, since larger budgets provide government agencies with more resources and personnel. We find no significant relationship between the size of the budget and the proportion of audits to the ministry that found irregularities.

Table A1. Budget size is not a predictor of irregularities found by auditors

Audits with irregularities	
Budget (log)	-0.01 (0.005)
Constant	0.44*** (0.10)
Observations	539
R <sup>2</sup>	0.003
F Statistic	1.59

*Note:* OLS regression with standard errors in parenthesis. Outcome is the proportion of audits to the ministry that found irregularities. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## 7. Placebo test with simulated anomaly scores

To rule out a spurious relationship, we re-run models from Table 1 with simulated anomaly scores randomly drawn from a gamma distribution (truncated between 0 and 1) using the moments of the original anomaly scores. We find no significant relationship.

Table A2. Placebo test with simulated anomaly scores

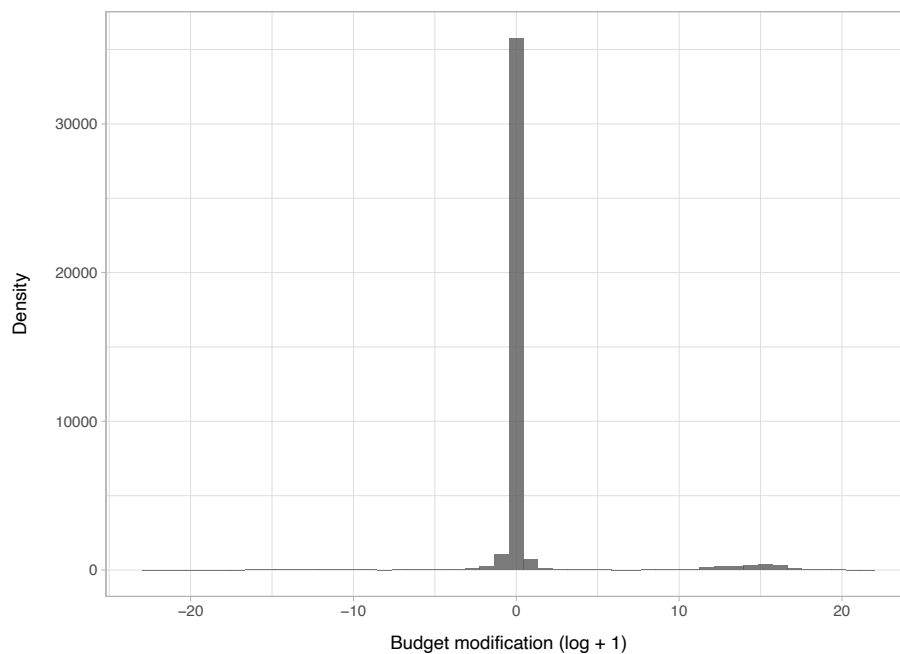
	Irregularities
Anomaly scores (simulated)	-0.09 (0.13)
Constant	0.15*** (0.03)
Observations	600
R <sup>2</sup>	0.001
F Statistic	0.40
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

*Note:* OLS regression with standard errors in parenthesis. Outcome is the total amount of irregularities found in audits to the ministry. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## 8. Details on the simulations

The goal of the simulation was recreating the distribution using information from the original moments of the original data on budget modifications. These entailed several challenges. Figure A6 showcases a typical example of a distribution, taking the logarithm to more clearly illustrate. In a typical distribution, most cases are zeroes, and there are extreme positive and negative values (more often extreme positive values).

Figure A6. Typical distribution of budget modifications



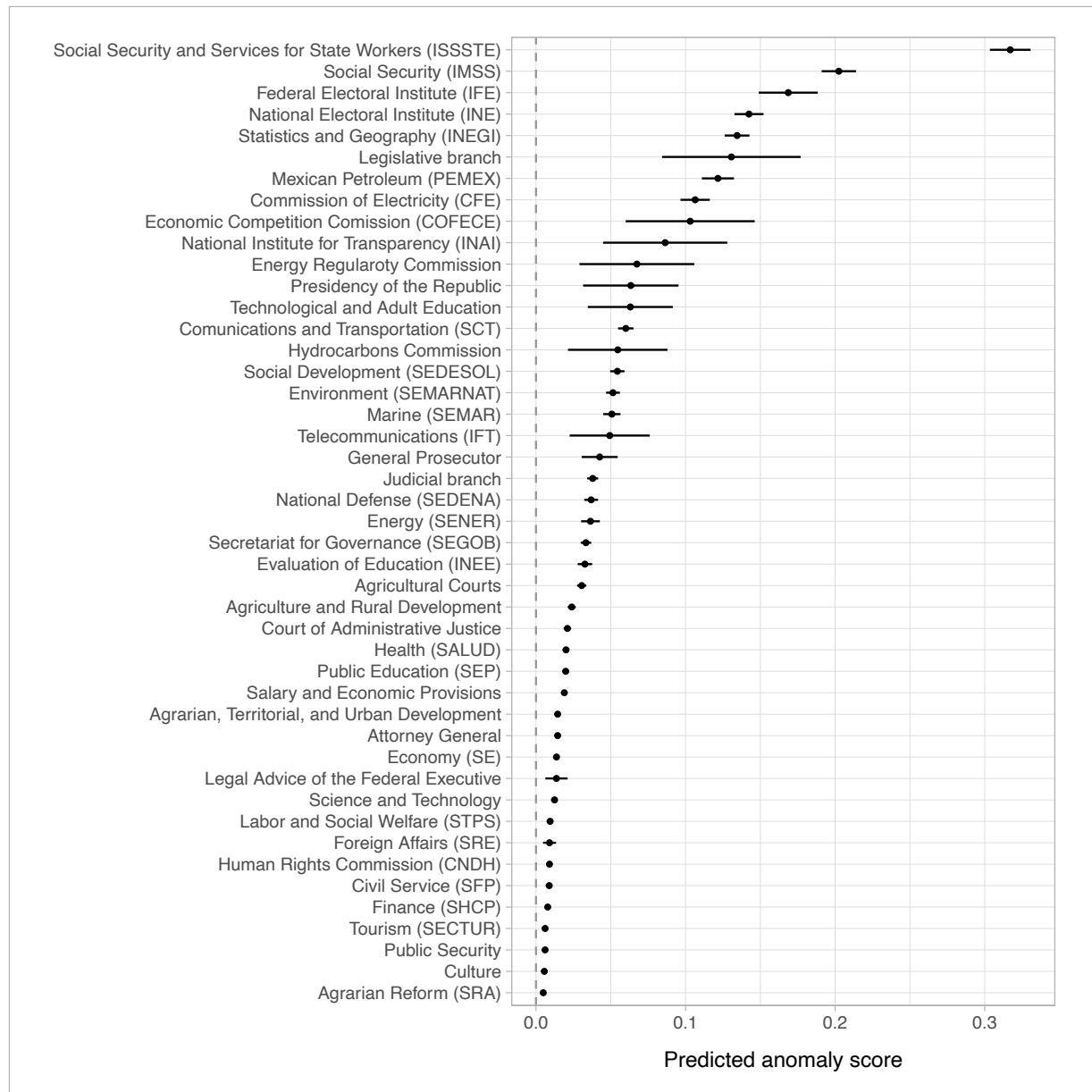
To create the synthetic data, we followed three steps:

1. We drew observations from a gamma distribution using the moments of the original variables.
2. We obtained the proportion of zeroes and negative values of the original variable.
3. Afterwards, we randomly multiplied random observations times zero or -1 so as to get a proportion of zeroes and observations with negative values that was close to the original distribution.

## 9. Ranking of anomaly scores by government ministry

Figure A7 plots the predicted values of a beta regression with ministry dummy variables as predictors of anomaly scores.

Figure A7. Predicted anomaly scores for ministries in the Mexican government



*Note:* Beta regression predicting anomaly scores with ministry dummy variables. Unit of analysis is a ministry-year.



## 10. Collection of news articles on malfeasance scandals

Table A3. Collection of news articles on malfeasance scandals

Social Security (ISSSTE/IMSS)
<p><b>Summary:</b></p> <ul style="list-style-type: none"> <li>- Institutions with most corruption complaints filed to the SFP were IMSS and ISSSTE.</li> <li>- Investigations identified nearly 23 thousand fake invoices in the IMSS and ISSSTE between 2014 and 2019, which led to the diversion of 4.1 billion pesos.</li> <li>- A federal judge linked public officials of the ISSSTE to a fraud for more than 831 million pesos, related to the acquisition of equipment medical and healing material.</li> <li>- There are 300 labor lawsuits in progress for fraud in ISSSTE pensions.</li> <li>- Fraud scandals related to fake widows collecting pension payments.</li> <li>- A federal judge linked the legal director of ISSSTE and seven more former officials for an alleged fraud of more than 831 million pesos.</li> <li>- The Specialized Prosecutor for Combating Corruption have detected hundreds of cases in which ISSSTE employees colluded with law firms and members of Boards Conciliation and Arbitration to pay inflated pensions.</li> <li>- The general director of ISSSTE, Sebastián Lerdo de Tejada, revealed the existence of investigations for various acts of corruption incurred by former officials of this institution, totaling more than 43 billion pesos.</li> <li>- Opinion column discussing the endless irregularities in ISSSTE purchases. Purchases are often overpriced, and tenders are assigned to companies without experience where there are connections to family members of public officials.</li> <li>- The Attorney General of the Republic filed an accusation against the current legal director of ISSSTE and seven other officials for alleged acts of corruption that caused the loss of more than 800 million pesos.</li> <li>- One of the main pharmaceutical groups in the country returned to ISSSTE more than 950 million pesos that it had obtained in an alleged fraudulent operation. This, while the criminal investigation continues against eight former officials, including the former lawyer of the institute itself, for corruption crimes.</li> </ul> <p><b>Sources:</b></p> <p><a href="https://www.eleconomista.com.mx/politica/Desestiman-mas-del-50-de-las-denuncias-por-corrupcion-20221222-0122.html">https://www.eleconomista.com.mx/politica/Desestiman-mas-del-50-de-las-denuncias-por-corrupcion-20221222-0122.html</a></p> <p><a href="https://contralacorrupcion.mx/anuario-de-la-corrupcion-2020/">https://contralacorrupcion.mx/anuario-de-la-corrupcion-2020/</a></p> <p><a href="https://www.jornada.com.mx/notas/2022/04/29/politica/fraude-al-issste-por-831-millones/">https://www.jornada.com.mx/notas/2022/04/29/politica/fraude-al-issste-por-831-millones/</a></p>

<https://www.jornada.com.mx/notas/2022/11/03/sociedad/en-curso-300-juicios-laborales-por-fraude-en-pensiones-del-issste/>

<https://www.elfinanciero.com.mx/nacional/2022/12/06/viudas-falsas-en-imss-e-issste-asi-es-el-fraude-que-hacen-en-pensiones/>

<https://aristeginoticias.com/2804/mexico/fgr-acusa-a-ex-funcionarios-del-issste-por-presunto-fraude-por-800-mdp/>

<https://www.jornada.com.mx/2022/06/05/politica/003n1pol>

<http://comunicacion.senado.gob.mx/index.php/periodo-ordinario/boletines/9154-boletin-405-investigaciones-por-corrupcion-en-el-issste-suman-mas-de-43-mmmdp-lerdo-de-tejada.html>

<https://www.eleconomista.com.mx/opinion/Mas-compras-sospechosas-en-el-ISSSTE-20230118-0003.html>

<https://www.animalpolitico.com/politica/fgr-acusa-funcionarios-issste-regalar-dinero-farmaceutica>

<https://www.animalpolitico.com/sociedad/empresa-implicada-fraude-issste-regresa-950-mdp>

#### Federal Electricity Commission (CFE)

##### Summary:

- A US judge rejected requests by two former executives of the CFE company to stop a case against them for alleged corruption in contracts worth billions of dollars.
- After carrying out 12 performance audits, the ASF concluded that the theft of electricity and the overdue portfolio during 2018 deteriorated the finances of the CFE. The ASF specified that energy theft had an estimated cost of 13 percent of the CFE's total income.
- The Federal Public Ministry obtained 18 arrest warrants against public servants of the CFE for an alleged fraud of 280 million pesos.
- Opinion column discussing corruption and inefficiency in the CFE.
- In the last 15 years, more than 40 CFE officials have received fines from the SFP for participating in acts of corruption. These add up to more than 80 million dollars.
- In 2019, the Campeche City Council sued the CFE with the intention that they make the return of two payments: one for two million and another for 834 thousand pesos.

##### Sources:

<https://elpais.com/mexico/2023-04-08/la-justicia-de-ee-uu-avanza-en-el-caso-de-dos-exejecutivos-de-cfe-acusados-de-corrupcion-amiguismo-y-conflicto-de-intereses.html>

<https://www.elsoldemexico.com.mx/mexico/sociedad/robo-de-luz-con-diablitos-afecto-a-cfe-asf-4403684.html>

<https://www.jornada.com.mx/2007/07/31/index.php?section=politica&article=016n1pol>

<https://www.elfinanciero.com.mx/opinion/lourdes-mendoza/2022/11/30/saqueo-corrupcion-y-crimen-en-cfe/>

chrome-

extension://efaidnbmnnnibpcajpcgiclfndmkaj/https://www.wilsoncenter.org/sites/default/files/media/documents/publication/corrupcion\_en\_el\_sector\_energetico\_mexicano\_propuestas\_y\_recomendaciones.pdf

<https://www.reporteindigo.com/reporte/desvios-millonarios-de-cfe-en-campeche-devolucion-pagos/>

#### National Guard

##### Summary:

- The National Guard has received large budgets while still relying on the Ministry of Defense for most of their expenditures. The Secretary of Defense paid for the construction of their military quarters, furnishing, equipment, and vehicles. Where is the National Guard's budget going to then? Their budget is basically used for basic services and duplicates payments in salaries, since 59% of their members are also members of the military and still continue to receive their respective salaries.

##### Sources:

chrome-extension://efaidnbmnnnibpcajpcgiclfndmkaj/https://contralacorrupcion.mx/wp-content/uploads/2021/09/los-militares-y-la-guardia-nacional.pdf?utm\_source=site&utm\_medium=button&utm\_campaign=mcci

#### Legislative branch

##### Summary:

- Mexican legislators diverted 506 million pesos to 130 shell companies between 2013 and 2018.
- President Andrés Manuel López Obrador considered that those who oppose the disappearance of the trusts do so because they endorse the acts of corruption that prevailed around these budgetary mechanisms.
- The Senate approved the extinction of 109 legislative trusts, which represent more than 68 billion pesos. The proposal is aimed at eradicating the opacity and corruption in trusts.

**Sources:**

<https://elpais.com/mexico/2021-01-27/los-diputados-mexicanos-desviaron-506-millones-de-pesos-a-130-empresas-fantasma-entre-2013-y-2018.html>

<https://www.capital21.cdmx.gob.mx/noticias/?p=3930>

<http://comunicacion.senado.gob.mx/index.php/informacion/boletines/49479-aprueba-senado-en-lo-general-extincion-de-109-fideicomisos.html>

<https://piedepagina.mx/detallan-opacidad-y-presunta-corrupcion-en-uso-de-fideicomisos/>

### Prevention and Social Rehabilitation

**Summary:**

- 61 orders of apprehension were emitted towards public officials linked to General Garcia Luna (former Secretary of Public Safety) for charges of embezzlement of public funds and ties to Narcos. They were accused of sacking public funds from the prevention and social rehabilitation administrative entity during the Calderon presidency.
- Public official, former employee of Garcia Luna, is being investigated for deviating 2,678 millions of pesos and 74 million USD. The Unit of Financial Intelligence detected transfers from the decentralized organ of prevention and social rehabilitation.

**Sources:**

<https://www.eleconomista.com.mx/politica/Van-por-Garcia-Luna-y-60-mas-por-peculado-20230522-0107.html>

<https://www.reforma.com/pide-general-amparo-por-desvios-en-gestion-de-garcia-luna/ar2646573>

<https://contralacorrupcion.mx/genaro-garcia-luna-nunvav-vanesa-pedraza/>

### Social Development (SEDESOL)

**Summary:**

- An investigation coined “La Estafa Maestra” found that SEDESOL participated in a fraud for 2.2 million pesos.
- The head of SEDESOL reported that there are over 500 criminal complaints filed against officials of the agency are in the process of being investigated for alleged acts of corruption.
- The ASF detected embezzlement of 1.3 millions of pesos, with payments to shell companies made by SEDESOL and SEDATU.

- All the agreements that SEDESOL signed with the autonomous universities of Morelos and the State of Mexico led to different operations to divert resources.
- For the Attorney General of the Republic the diversion of resources carried out SEDESOL and SEDATU, during the management of Rosario Robles between 2012-2018 constitute one of the most scandalous cases of corruption and patrimonial damage to the country, and up to now, the amount looted exceeds five billion pesos and there are 24 federal criminal cases involving more than 50 people.

**Sources:**

<https://www.animalpolitico.com/estafa-maestra/sedesol-donde-quedo-dinero-pobres.html>

<https://www.eleconomista.com.mx/politica/En-Sedesol-hay-500-denuncias-por-corrupcion-20181024-0173.html>

<https://elceo.com/politica/desvio-de-recursos-y-contratistas-fantasma-los-casos-de-presunta-corrupcion-en-sedesol/>

<https://contralacorrupcion.mx/web/estafamaestra/sedesol-donde-quedo-dinero-pobres.html>

<https://www.jornada.com.mx/notas/2023/02/25/politica/escandaloso-caso-de-corrupcion-de-robles-en-sedesol-y-sedatu-fgr/>

## Mexican Petroleum (PEMEX)

**Summary:**

- Several ex-Pemex directors under investigation or prison for cases of fraud and corruption.
- From 2003 to 2012, Reuters identified more than 100 contracts with an approximate value of 117 billion dollars (8% of Pemex's total income) that, according to the ASF, presented serious irregularities. These included contract overvaluations, phantom employees, payments for services not received, conflicts of interest, etc.
- Pemex, currently under investigation for the alleged bribes that some of its officials received from the Brazilian construction company Odebrecht, is the instance of the Mexican Government that concentrates the highest sanctions for corruption cases. Fines imposed to Pemex in the last 20 years amount to 22,119 million pesos.
- Opinion piece discussing rampant corruption in Pemex.
- The repair of damages offered by the former director of Pemex, Emilio Lozoya, for irregular purchases and bribes could double and reach the figure of 20 million dollars.
- Opinion piece discussing robbery and corruption in Pemex warehouses.

**Sources:**

<https://expansion.mx/empresas/2021/11/04/directores-pemex-en-carcel>

chrome-

extension://efaidnbmnnnibpcajpcgglefindmkaj/https://www.wilsoncenter.org/sites/default/files/media/documents/publication/corrupcion\_en\_el\_sector\_energetico\_mexicano\_propuestas\_y\_recomendaciones.pdf

<https://www.elfinanciero.com.mx/nacional/2021/11/04/son-ya-4-exdirectores-de-pemex-enjuiciados/>

[https://elpais.com/internacional/2017/03/22/mexico/1490222209\\_910802.html](https://elpais.com/internacional/2017/03/22/mexico/1490222209_910802.html)

<https://forojuridico.mx/pemex-saqueo-historico-por-corrupcion-que-aun-se-puede-rescatar/>

<https://www.telesurtv.net/news/director-pemex-mexico-emilio-lozoya-corrupcion-gobierno-20230119-0012.html>

<https://www.elfinanciero.com.mx/opinion/lourdes-mendoza/2023/01/23/corrupcion-en-almacenes-y-talleres-centrales-de-pemex/>

### Communications and Transportation (SCT)

#### Summary:

- The SCT filed a criminal complaint with the State Attorney General's Office for the potential crimes of bribery, embezzlement, and corruption in construction contracts filed with previous administrations.
- SCT found hundreds of alleged irregularities in public procurement from the past administration and about 482 observations still to be addressed.
- Opinion piece discussing the "highway mafia" that is the Zacatecas SCT.
- Simulated contracts with shell companies amount to 190 million pesos.
- Opinion piece discussing corruption in SCT public procurement and simulated contracts.
- Industrial construction entrepreneurs from various states of the country denounced "rigged bids" and "charge of kickbacks up to 20 percent" for the award of public works, mainly in the states of San Luis Potosí, Veracruz, Puebla, Tlaxcala and Jalisco.
- The SCT allocated 13 billion pesos to build roads in the poorest areas of the country, but there were not even enough supervisors to verify that the 553 works carried out in 2016 complied with the design and materials provided. Audits detected many irregularities.

#### Sources:

<https://www.eluniversal.com.mx/estados/sct-de-san-luis-potosi-presenta-denuncia-penal-por-diversos-delitos-en-administraciones-pasadas/>

<https://revistamagazine.com/camiones/sct-pretende-ser-una-zona-libre-de-corrupcion/>

<https://www.elsoldezacatecas.com.mx/analisis/la-corrupcion-en-comunicaciones-y-transportes-9010194.html>

<https://dplnews.com/la-huella-de-corrupcion-de-jimenez-espriu-en-sct/>

<https://www.eluniversal.com.mx/opinion/mario-maldonado/en-la-sct-huele-corrupcion>

<https://www.elfinanciero.com.mx/nacional/se-va-jimenez-espriu-y-surgen-denuncias-de-corrupcion-en-la-sct/>

<https://www.animalpolitico.com/sociedad/sct-verificar-carreteras-corrupcion>

#### National Defense (SEDENA)

**Summary:**

- The former Secretary of National Defense, Salvador Cienfuegos, was arrested in Los Angeles and accused of drug trafficking and money laundering.
- The Mexican Army diverted 156 million dollars to shell companies between 2013 and 2019.

**Sources:**

<https://contralacorrupcion.mx/anuario-de-la-corrupcion-2020/>

<https://elpais.com/mexico/2020-08-25/el-ejercito-mexicano-desvio-156-millones-de-dolares-a-empresas-fantasma-entre-2013-y-2019.html>

#### National Electoral Institute (INE/IFE)

**Summary:**

- A citizen complaint filed yesterday before the Superior Audit Office of the Federation (ASF) demands a political trial against whoever is responsible for possible diversion of resources in the INE. Financial inconsistencies were detected in trusts. In addition, 11 direct awards were detected that did not have the required documentary justification.
- The Executive Director of Administration, recognized that the irregularities in 60 million pesos were to solve part of the past electoral process in the state of Guerrero, because the entity did not have the political, social and adequate security.
- A federal judge linked four INE officials to a process of improperly awarding food contract for events for 105 million pesos, of which they paid 15 million.

**Sources:**

<https://www.jornada.com.mx/notas/2022/04/16/politica/denuncia-ciudadana-ante-asf-exige-juicio-politico-por-desvio-de-recursos-en-el-ine/>

<https://www.jornada.com.mx/2009/09/21/opinion/009n1pol>

<https://www.sndigital.mx/mexico/20262-en-el-ine-no-hay-desvio-de-recursos-benito-nacif.html>

<https://lucsdelsiglo.com/2020/09/09/detectan-en-ine-desvio-de-15-mdp-nacional/>

#### National Institute of Statistics and Geography (INEGI)

**Summary:**

- The INEGI comptroller reported that in 2021 they received 3,000 citizen complaints for alleged irregularities within the institute, 1,100 liability procedures were processed and 110 observations were determined in the audits.
- INEGI diverted around 800 million pesos and violated the Budget and Public Expenditure Law. The ASF determined that there have been illegal acts in the agency's budget operation and asked to sanction public officials involved.
- Denounces of diversion of public resources in the INEGI for 178 million pesos.
- Operational workers of the INEGI 2020 census denounced diversion of resources in the management of the payroll and travel expenses of the pollsters.

**Sources:**

<https://www.heraldo.mx/pone-inegi-bajo-lupa-a-sus-colaboradores/>

<https://www.jornada.com.mx/2007/08/16/index.php?section=sociedad&article=042n2soc>

<https://vanguardia.com.mx/noticias/nacional/2841161-denuncian-desvio-de-recursos-publicos-en-el-inegi-por-178-millones-de-pesos-HVVG2841161>

<https://laopcion.com.mx/local/denuncian-desvio-de-recursos-en-la-nomina-de-inegi-20200324-265585.html>

**Tax Administration Service (SAT)****Summary:**

- Former PRI secretary arrested as part of an investigation into the triangulation of public money to campaigns. A protected witness stated that money had been diverted in six states, with shielding from SAT.
- Former officials of SAT were denounced before the Attorney General's Office for the crimes of embezzlement, bribery and improper use of powers. They gave a contract for more than one billion pesos to a company that was disabled.
- Corruption complaints filed against SAT increased by 96% in 2020.
- At least 11 officials of the SAT were denounced for acts of corruption and some of them received up to four years in prison.
- The more than 3,000 criminal complaints filed by SAT during the presidency of Andrés Manuel López Obrador to punish corrupt practices in the tax field include officials of the agency itself.



- In 2019, SAT ranked among the institutions with most complaints of corruption filed to the SFP. Out of 589 ongoing investigations of embezzlement, SAT was the institution with the largest number of investigations.
- The screenshot shows an add on twitter that encourages victims of corruption by SAT officials to submit a report to the anticorruption portal.

#### Sources:

<https://aristeginoticias.com/2112/mexico/senala-testigo-que-sat-participo-en-desvio-de-recursos-federales-a-campanas-del-pri/>

<https://www.jornada.com.mx/2019/02/18/politica/007n2pol>

<https://www.forbes.com.mx/economia-aumenta-96-las-denuncias-por-corrupcion-en-el-sat/>

<https://www.elcontribuyente.mx/2022/11/como-castigo-el-sat-a-sus-funcionarios-acusados-de-corrupcion/>

<https://www.reporteindigo.com/reporte/la-limpia-del-sat-contr-la-corrupeion-registro-denuncias-detenciones-abusos/>

<https://www.eleconomista.com.mx/politica/Registra-SFP-mas-de-200000-denuncias-por-anomalias-en-instituciones-20220804-0125.html>



### National Institute of Migration

**Summary:**

- Out of 589 ongoing investigations of embezzlement with public funds, the National Institute of Migration is the second agency with the largest number of investigations.

**Sources:**

<https://www.eleconomista.com.mx/politica/Registra-SFP-mas-de-200000-denuncias-por-anomalias-en-instituciones-20220804-0125.html>

### Instituto Politécnico Nacional (IPN)

**Summary:**

- IPN students recently held protests against corruption, scarcity of resources, and harassment in the institution.
- The FGR and the SFP are investigating alleged acts of corruption that include the diversion of 9 million pesos from the IPN treasury.
- A total of seven schools of the IPN are on strike, accusing alleged acts of corruption and complaints against teachers for mistreatment and inflexibility and for abuse of functions by management personnel.
- Principal and deputy principal have been accused of various crimes, including diversion of resources by appointing and occupying non-existent positions, abuse of duties, and even bribery against teachers. Officials created fake positions and third-party contracts, using those accounts to embezzle funds away from IPN.
- For decades, the IPN has simulated contracts and engaged in acts of corruption. The National Confederacy of Professionals and Youths (CONAPRO) legally accused the IPN.

**Sources:**

<https://elpais.com/mexico/2022-09-09/los-estudiantes-del-ipn-se-rebelan-frente-a-la-corrupcion-la-escasez-de-recursos-y-el-acoso.html>

<https://www.eleconomista.com.mx/arteseideas/Urgen-a-la-SCJN-intervenir-ante-presuntos-actos-de-corrupcion-en-el-IPN-20220711-0008.html>

<https://www.excelsior.com.mx/comunidad/siete-planteles-del-ipn-en-paro-denuncian-corrupcion/1538178>

<https://www.radioformula.com.mx/cdmx/2022/8/22/escandalo-en-el-poli-asi-es-la-supuesta-corrupcion-de-los-directivos-de-escuela-del-ipn-728651.html>

<https://www.milenio.com/politica/denuncian-supuestos-actos-corrupcion-ipn-conapro>

#### National Water Commission (CONAGUA)

**Summary:**

- CONAGUA dismisses more than 150 officials for alleged acts of corruption.
- In the midst of the controversies over alleged acts of corruption and the alleged delivery of concessions during the current administration, CONAGUA has added a series of personnel changes, from the owner to administrative workers.
- Report that compiles investigations and criticizes the arbitrary and politicized use of water. Mentions of fraud, black markets, and corruption in public procurement processes.
- Between 2016 and 2018, the ASF found that CONAGUA embezzled 270 million pesos through inflated prices in and cheap construction materials for tubing in rivers.

**Sources:**

<https://politica.expansion.mx/mexico/2022/04/21/conagua-cesa-a-mas-de-150-funcionarios-por-presuntos-actos-de-corrupcion>

<https://politica.expansion.mx/mexico/2022/11/11/corrupcion-en-la-conagua-esto-se-sabe>

<https://contralacorrupcion.mx/explotadores-agua-mexico/>

<https://www.milenio.com/politica/comunidad/fraude-conagua-mala-tuberia-provoco-inundacion-ecatepec>

#### Federal Police

**Summary:**

- The Federal police embezzled over 2 thousand 657 million pesos in an uninterrupted manner for 6 years (starting in 2013). Money disappeared, public contracts simulated, and prices for public security equipment inflated.
- In the second half of 2018, 40 million dollars were diverted from the budget of the Mexican Federal Police and towards a network of companies.
- In August 2020, arrest warrants were issued against 19 former officials of the Federal Police for alleged diversion of resources and overpriced purchases between 2012 and 2018.

- The Ministry of Public Security said in a report that it detected irregularities in the transfer of resources from the Federal Police. The alleged diversion network occurred through the purchase of planes, helicopters, and patrol cars at premium prices; also through simulation in acquisitions.

**Sources:**

<https://www.animalpolitico.com/2020/08/desvios-policia-federal-2-mil-650-mdp>

<https://elpais.com/mexico/2020-12-08/la-piramide-asi-desviaron-40-millones-de-dolares-del-presupuesto-de-la-policia-federal-en-los-ultimos-seis-meses-del-gobierno-de-pena-nieto.html>

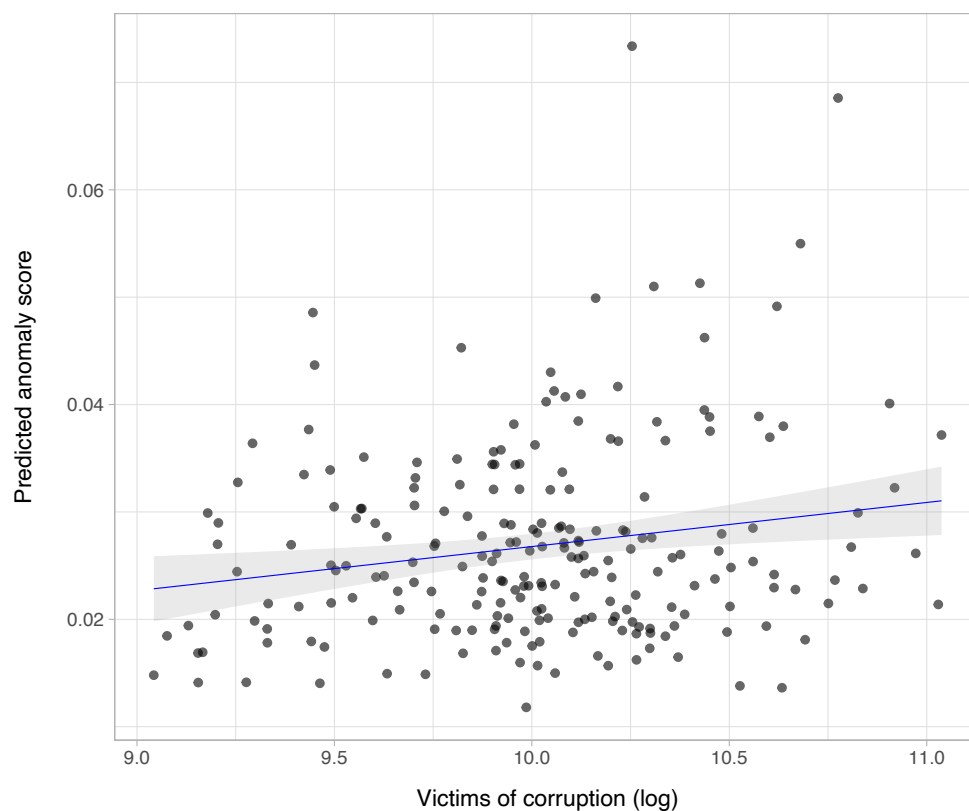
<https://www.eluniversal.com.mx/nacion/amparan-13-funcionarios-de-la-policia-federal-acusados-de-delincuencia-organizada-en-sexenio-de-eqn/>

<https://businessinsider.mx/detencion-eleuterio-enrique-perez-romero-miguel-angel-osorio-chong-desvio-de-recursos-policia-federal/>

## 11. State-level perceptions of corruption

The National Office of Statistics and Geography (INEGI) carries out a survey of government quality and impact (ENCIG) every two years. One of the questions includes the number of victims of corruption in each state. This measure captures the number of respondents in the that reported to experience corruption when interacting with a public official. Below, Figure A8 shows the scatterplot with the positive relationship ( $corr = 0.18$ ) and Table A3 reports the regression results.

Figure A8. Victims of corruption and predicted anomaly scores for Mexican states



*Note:* Y-axis shows the predicted anomaly score, obtained through a beta regression predicting anomaly scores with state-year dummy variables. X-axis is the number of respondents in the official survey of government quality and impact (ENCIG) that reported to experience corruption when interacting with a public official (log). Survey data available for 2019, 2017, 2015, and 2013. Data for missing years was imputed with the average of the previous and upcoming year.

Table A4. Victims of corruption and predicted anomaly scores for Mexican states

	Victims of corruption
Predicted anomaly score	8.64*** (3.02)
Constant	9.77*** (0.09)
Observations	224
R <sup>2</sup>	0.04
F Statistic	8.17***

*Note:* OLS regression with standard errors in parenthesis. Unit of analysis is a state-year. Outcome is the log of victims of corruption in a state-year. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## 12. Top anomalous agencies and categories of expenditure

Tables A5-A6 list the agencies with highest anomaly scores, along with their most anomalous expenditure categories. For both, anomaly scores were first calculated at the agency-year level and then ranked based on that number. In order to find the most anomalous expenditure categories, isolation forest was then run within an agency-year and top categories were then listed. Table A5 shows results at the agency-year level and Table A6 excludes cases in Mexico City. Since Mexico City has a considerably larger population, budgets for those agencies are larger. Even after normalizing, budget modifications in Mexico City stand out.

Table A5. Top anomalous agencies and expenditure categories

Agencies		Expenditure categories
Social Services of State Workers (ISSSTE, Mexico City 2014)	Pensions and retirements   Salaries	
Federal Police (2013)	Bonus for special services   Salaries   Expenditure on public security	
Social Services of State Workers (ISSSTE, Mexico City 2018)	Medicines and pharmaceutical products   Salaries   Medical equipment	
Supreme Court (2012)	Salaries   Bonuses   Collective contracts	
Social Services of State Workers (ISSSTE, Mexico City 2019)	Third-party services   Medicines and pharmaceutical products   Salaries	
Council of the Federal Judicature (2012)	Accreditation for teaching   Salaries	
Social Services of State Workers (ISSSTE, Mexico City 2015)	Medical equipment   Other benefits   Pensions and retirements	
Social Services of State Workers (ISSSTE, Mexico City 2017)	Federal transfers   Distributions for third-parties   Productivity subsidies	
Electoral Tribunal of the Federal Judiciary (2012)	Salaries   Bonuses   Other benefits	
Supreme Auditor of the Federation (ASF, 2012)	Bonuses   Salaries	
Social Services of State Workers (ISSSTE, Mexico City 2016)	Federal transfers   Other benefits   Fees for funds and insurance	
Tax Administration Services (SAT, 2014)	Transfers to public trusts   Maintenance of property   Travel payments for public officials	
National Polytechnic Institute (IPN, 2011)	Third-party services   Basic services   Surveillance services	
Federal Court of Fiscal and Administrative Justice (Mexico City 2012)	Salaries   Bonuses   Furniture	
Senate (2012)	Funds to parliamentary groups   Salaries   Furniture	
National Guard (2020)	Bonus for special services   Salaries   Fuels and additives	
Prevention and Social Rehabilitation (2013)	Integral services   Bonus for special services   Salaries	
Tax Administration Services (SAT, 2015)	Transfers to public trusts   Maintenance of property   Travel payments for public officials	
Tax Administration Services (SAT, 2012)	Transfers to public trusts   Travel payments for public officials   Basic services	
Tax Administration Services (SAT, 2013)	Transfers to public trusts   Distributions for resolutions   Studies and research	
Institute of Geography and Statistics (INEGI, Aguascalientes 2012)	Bonuses   Salaries	
Tax Administration Services (SAT, 2011)	Transfers to public trusts   Payroll tax   Lease of equipment and machinery	
Social Services of State Workers (ISSSTE, Mexico City 2020)	Medical equipment   Salaries   Bonuses	
Federal Police (2019)	Bonus for special services   Salaries	
National Water Commission (CONAGUA, Mexico City 2016)	Transfers to public mandates   Subsidies   Vehicles (exclusive for natural disasters)	

*Note:* Top 25 anomalous agencies and their most anomalous expenditure categories, ranked from highest to lowest.



Table A6. Top anomalous agencies (excluding Mexico City)

Agencies
Council of the Federal Judiciary (2012)
National Polytechnic Institute (IPN, 2011)
Institute of Geography and Statistics (INEGI, Aguascalientes 2012)
Tax Administration Services (2011)
Federal Police (2011)
National Institute for Adult Education (2010)
National Commission for the Development of Indigenous Towns (2011)
National System for the Comprehensive Development of Families (2011)
Mexican Petroleum (PEMEX, Veracruz 2017)
Mexican Petroleum (PEMEX, Veracruz 2016)
National Forest Commission (2010)
<i>Oportunidades</i> Social Program (2011)
National Institute for Adult Education (2011)
National Banking Commission (2011)
Social Services of State Workers (ISSSTE, Veracruz 2014)
Institute of Geography and Statistics (INEGI, Aguascalientes, 2013)
<i>Oportunidades</i> Social Program (2010)
National Electoral Institute District Boards (Jalisco 2012)
Mexican Petroleum (PEMEX, Tamaulipas 2017)
Mexican Petroleum (PEMEX, Tamaulipas 2016)
IFE District Boards (Veracruz 2012)
IFE District Boards (State of Mexico 2012)
Federal Commission of Electricity (San Luis Potosi 2019)
Federal Police (2010)
Institute of Geography and Statistics (INEGI, Aguascalientes 2014)

*Note:* Top 25 agencies with the largest anomaly scores, ranked from highest to lowest (excluding Mexico City).