Institutional persistence and the sanctioning gap: A longitudinal archival analysis of Zambia's domestic resource mobilization (2003–2023)

Isaac Munjunga

munjunga@gmail.com

https://orcid.org/0009- 0001-2120-6863

University of Zambia

https://doi.org/10.51867/ajernet.7.3.109

ABSTRACT

Zambia's fiscal landscape shows a recurring misalignment between formal institutional design and operational outcomes. Despite successive administrative and digital reforms, identical financial irregularities reappear annually since 2003. This paper investigates the "persistence paradox." This is a phenomenon wherein the same fiscal irregularities and institutional failures repeatedly reoccur and grow over time and remain fully unaffected by subsequent policy improvements or technological upgrades. These include revenue leakage, procurement fraud, and internal control failures that recur across political administrations and technological upgrades. A longitudinal archival document analysis of 21 consecutive Auditor General's reports (2003 - 2023) was conducted. Automated text-mining in R extracted monetary values of fiscal irregularities; deductive thematic coding grounded in New Institutional Economics (NIE) and Good Governance Theory (GGT) categorized recurring failure modes. Uncollected taxes increased from K2.56 billion (2005) to K97.86 billion (2022). Procurement fraud via "unvouched expenditure" and systemic Value Added Tax (VAT) refund delays (averaging >845 days) are structural features. A critical "Sanctioning Gap," which is a structural void between the discovery of a financial irregularity by audit institutions and the actual enforcement of a penalty, ultimately leads to institutionalizing a culture of impunity. In the end, formal disciplinary codes are systematically undermined by informal impunity norms, high enforcement transaction costs, and isomorphic mimicry of digital systems. To this effect, Zambia's 9th National Development Plan (9NDP) must move beyond ceremonial digital adoption and ensure there is elimination of the Sanctioning Gap. It is recommended from this study that: (i) Zambia should legislate automated sanctioning mechanisms; (ii) enforce end-to-end digital integration between Zambia Revenue Authority (ZRA), line ministries, and Treasury; (iii) adopt performance contracts based on a "Net Audit Compliance Rate"; and (iv) launch a public accountability tracker to close the fiscal social contract deficit.

Keywords: Auditor General, Archival Document Analysis, Domestic Resource Mobilization, Institutional Failure, Public Finance Management, Zambia
JEL Codes: H30, O17, D73

I. INTRODUCTION

The efficiency of Domestic Resource Mobilization (DRM) is the primary determinant of a developing nation's path toward fiscal sovereignty (Ayana et al., 2023; Junquera-Varela et al., 2017; Oyinlola et al., 2020). In Sub-Saharan Africa, Zambia represents a critical case of the "development paradox" where a resource-rich nation struggling with chronic revenue underperformance and debt (Siwale & Gong, 2023). In November 2020, the culmination of structural fiscal deficits and escalating debt servicing costs materialized when Zambia became the first African nation in the pandemic era to default on its sovereign obligations, triggered by a missed US$42.5 million coupon payment (Mbewe et al., 2024) (Williams, 2020). This resulted in immediate credit downgrades to "Restricted Default" and "Selective Default" by Fitch and S&P Global Ratings, respectively, effectively locking Zambia out of international capital markets.

This default stems from a failure to raise enough revenue to be able to repay the debt and invest in economic development. Revenue sources and collection have not helped as the finances have not grown. The revenue yield is lower than needed resources. Despite the undertaking of reforms and sequential National Development Plans (NDPs) that have been implemented since independence (Munjunga et al., Integrating a medium-term revenue strategy into Zambia's 9th National Development Plan., 2026), Zambia suffers from a revenue gap, with recent estimates of the tax gap to be between 47 and 56 percent (Adu-Ababio et al., 2023; Lungu & Mungule, 2023). It is hard for Zambia to raise enough DRM to fund its operations and invest in development. Recent empirical work (Munjunga et al., Institutional quality, domestic resource mobilization, and economic growth in Zambia: A quantitative ARDL analysis (1990–2021), 2026) show rigidities in tax collection that negatively affects Gross Domestic Product (GDP) Per Capita growth. Even though non-tax collections have a positive contribution from the empirical study, its effect on economic growth is not very pronounced.

Licensed Under Creative Commons Attribution (CC BY-NC)

To accurately trace the root of this empirical problem, a longitudinal methodology is required. In contrast to econometric growth models that typically treat "institutional quality" as a latent or static variable, therefore, this study utilizes the empirical findings of the Auditor General (AG) reports to provide a granular, longitudinal assessment of institutional performance *in situ*. By treating "time as data," over two decades of AG reports is analysed to understand why specific irregularities, ballooning tax refund arrears, procurement fraud, and internal control failures that have been persist across different political administrations and regimes (Munjunga et al., Integrating a medium-term revenue strategy into Zambia's 9th National Development Plan., 2026). It is contended that Zambia's fiscal management is compromised by a deeply rooted problem that transcends political cycles, requiring a fundamental separation of accountability mechanisms from political discretion.

Zambia's experience is not isolated. It mirrors findings from other Sub-Sahara African countries (Sikhosana et al., 2025; Quashie, 2019). However, there is limited literature on Zambia that addresses fiscal failures by taking an in depth look at the Auditor General's official reports. Such an undertaking would initiate measures that would remedy the situation and help the country benefit from revenue collection. This gap will be filled by this study, by analysing Zambia's 21-year longitudinal records. These reveal an unusually high degree of institutional isomorphism (DiMaggio & Powell, 2015), where digital reforms are adopted ceremonially without altering of enforcement routines within the institutions.

1.1 Statement of the Problem

Many reforms are undertaken in Zambia with regard to Public Financial Management (PFM), it is expected that a lot of controls have been put in revenue management. Ideally Zambia should have functioning systems as a result of this with formal Public Financial Management (PFM) laws, tax administration protocols, and digital safeguards which should include the Integrated Financial Management Information System (IFMIS), Tax Online, and the Payroll Management and Establishment Control (PMEC) system. These are supposed to function collectively to ensure efficient and transparent DRM, eliminate off-budget leakages, and enforce administrative sanctions against offenders in order to maximize net revenue for funding and national development. In reality, Zambia's formal institutional design is profoundly decoupled from operational outcomes. With reports of "unvouched expenditures," uncollected mineral, uncollected ground rents, "ghost" workers, and chronic VAT refund delays that recur and escalate annually. With the failure to penalize Category B financial offenses which institutionalizes a culture of impunity, deprives the Treasury of mobilized resources, erodes the Net Audit Compliance Rate, and forces the state into capital starvation. Consequently, this study operationalizes "time as data" across the 2003 - 2023 Auditor General reports to empirically diagnose the Sanctioning Gap, explain the political economy mechanisms driving isomorphic mimicry, and formulate enforcement-centred policy interventions. In so doing, this study is aimed at proposing interventions that will remedy this, and propose measures that can be taken to ensure an end to this for Zambia's 9th National Development Plan (9NDP).

1.2 Research Objectives

From the above, the general objective of this study using the Auditor General reports from 2003 - 2023, is to investigate the longitudinal trajectory, structural bottlenecks or failures, and institutional mechanisms that lead to fiscal leakage in Zambia across the 21-year period. From this general objective, specific objectives are as follows:

i. To quantify the longitudinal scale and frequency of recurring fiscal irregularities: This will specifically look at uncollected revenue arrears, unvouched expenditures, and payroll fraud - from 2003 to 2023 using automated regex text-mining.
ii. To evaluate the impact of digital PFM interventions such as IFMIS, Tax Online and Payroll Management and Establishment Control (PMEC) on operational compliance, analysing for isomorphic mimicry by comparing the volume of audit citations before and after system deployment.
iii. To identify the institutional and political economy mechanics of the "Sanctioning Gap", by examining how administrative friction, procedural opacity, and elite capture prevent the enforcement of formal disciplinary codes.
iv. To operationalize and apply the "Net Audit Compliance Rate" as a structural diagnostic metric to evaluate the widening divergence between gross revenue collection and net retained public resources.
v. To formulate actionable, enforcement-centred policy recommendations for Zambia's 9th National Development Plan (9NDP), focusing on automated strict-liability sanctions and independent fiscal oversight.

II. LITERATURE REVIEW

Licensed Under Creative Commons Attribution (CC BY-NC)

2.1 Theoretical Review

The *Persistence Paradox* is the phenomenon where identical fiscal irregularities and institutional failures continuously recur and escalate over time, remaining completely unaffected by sequential policy reforms or technological enhancements and upgrades. It will be noted that fiscal irregularities escalate despite the introduction of modern digital reforms such as introduction of Integrated Financial Management Information System (IFMIS). This paradox is sustained by the *Sanctioning Gap*, which is a gap between the discovery of a financial irregularity by audit institutions and the actual enforcement of a penalty. A big gap would effectively institutionalize a culture of impunity.

In order to understand the *Persistence Paradox*, it is necessary to examine how institutional change actually occurs or fails to occur within national bureaucracies or entities. According to Mahoney and Thelen's historical institutionalism (2010), institutional change is rarely an abrupt and complete overhaul; rather, it often proceeds incrementally in processes like layering through adding new rules on top of old ones or conversion by redeploying existing rules for new purposes. Over time, a certain form of handling things will be developed and it may persist. In the Zambian fiscal context, digital reforms such as the IFMIS represent layering. A system that has been brought in to solve several issues that have been observed. Therefore, new digital constraints were layered over existing administrative hierarchies. However, because the underlying informal rules were not displaced and completely taken out, the system settled into a state of stasis. North (1990) attributes this stasis to the "increasing returns" of institutional matrices: once corrupt networks adapt to a set of rules that may even be dysfunctional ones, the transaction costs of transitioning to a genuinely transparent system threaten their accumulated political and financial capital gains. Without a critical exogenous shock coming in or a binding enforcement mechanism, such as an Independent Fiscal Tribunal, endogenous institutional change is effectively blocked by the beneficiaries of the Sanctioning Gap.

2.1.1 New Institutional Economics and Good Governance Theory in DRM

This study integrates New Institutional Economics (NIE) and Good Governance Theory (GGT) in explaining national economic development as institutional quality help with good governance initiatives. Institutions are the "rules of the game" shaping economic incentives (Acemoglu & Robinson, 2012; North, 1990). Formal and informal restrictions that stop extractive behaviour are essential to the effectiveness of any revenue system (Lungu & Mungule, 2023; Nikiema & Zore, 2025). Persistent irregularities in AG reports therefore, suggest an "extractive" institutional arrangement. According to GGT's normative framework (Apaza, 2009; IMF, 2023), sustainable development requires accountability, transparency, and the rule of law. DRM is a social compact and therefore when government does not enforce accountability for unpaid taxes, questionable collection, or procurement fraud, DRM legitimacy is compromised. For example, when VAT refund procedures fail economic activity is distorted (Lindunda et al., 2024).

NIE and GGT provide robust analytical and explanatory power in many studies. However, its application in Sub-Sahara African context sometimes faces valid scholarly critique, this study acknowledges and also addresses these. Scholars including Chang (2011)argue that NIE can be overly functionalist, sometimes assuming that institutions always evolve efficiently to reduce transaction costs. Further, NIE is said to frequently underestimate the role of colonial legacies and entrenched power asymmetries in shaping fiscal regimes. This study addresses this critique by demonstrating that the Zambian fiscal matrix is not evolving toward efficiency; rather, rational actors are utilizing administrative friction to deliberately maintain inefficiency and extract rents that they benefit from. The inefficiencies, therefore, are due to vested interests.

GGT is also frequently criticized for promoting a Western-centric, "one-size-fits-all" institutional template. Grindle (2004) argued well for "Good Enough Governance," positing that developing nations cannot logically be expected to achieve perfect transparency and rule of law simultaneously; and that pushing unrealistic GGT agendas often leads to institutional overload. Similarly, Andrews (2013) critiques GGT for incentivizing the exact isomorphic mimicry documented in this study. By shifting the focus from the mere adoption of formal laws to the actual rate of Net Audit Compliance, this study adapts GGT from an unrealistic normative checklist into a localized, measurable assessment of state capability. Therefore, NIE and GGT are well suited to handle this objectively.

2.1.2 The Political Economy of Revenue Collection in Africa

Fiscal leakage has negative effects on an economy as the revenue that has been collected, cannot be used in financing the country. Persistence of fiscal leakage has a lot of negative effects on the economy, not only on failure to fund but also on the institutional degradation as a result of the leakage. In understanding the persistence of fiscal leakage requires examining the political economy of revenue collection. This is more so that revenue collection is viewed not merely as an administrative task, but as a deep political bargaining process. This political aspect has effects on national building.

Taxation, State-Building, and the Fiscal Contract: A lot of literature indicate that taxation helps build the country and to provide public services. However, it should be noted that this is a fiscal contract. The foundational work of Mick Moore (2004) emphasizes that the manner in which a state raises its revenues fundamentally shapes its institutional

Licensed Under Creative Commons Attribution (CC BY-NC)

governance. Moore argues that for a "governance dividend" of taxation, wherein states that rely on broad-based, citizen-derived taxes are forced into a reciprocal fiscal social contract, leading to higher accountability and stronger democratic institutions. This is common in industrialized countries as they have succeeded in creating high levels of both capacity and consent (Brautigam et al., 2008). However, in resource-rich countries in Africa, like Zambia, the reliance on extractive rents like royalties from copper, severs this accountability link. When the country does not depend on the majority of its citizens for revenue, elites have little incentive to build a transparent, rule-bound administrative apparatus. Consequently, the fiscal social contract breaks down, driving widespread tax evasion and justifying the informal decoupling of formal rules from actual enforcement. This leads to failures in state-building.

Elite Capture and Administrative Corruption: Odd-Helge Fjeldstad's (2014) research on tax reform in Sub-Saharan Africa provides the operational context for the "Sanctioning Gap." Fjeldstad (2001, 2014) shows that tax evasion and revenue leakage are rarely as a result of lack of technical knowledge; rather, they are a product of deliberate collusion between political elites, technical elites, wealthy taxpayers, and tax administrators. Fjeldstad observes that powerful actors often exploit "reforms" to solidify their rent-seeking networks. When new rules and procedures are introduced, the administrative networks of officers adapt and utilize their discretionary power to bypass controls and protect their patronage systems. This theoretically validates the Elite Capture finding that uncollected tax arrears, exports without paying taxes and procurement fraud are sustained by elites who actively manipulate administrative procedures and shield themselves from sanctions in Zambia.

2.1.3 Mechanisms of Institutional Change and Stasis

To understand the "Persistence Paradox," it is necessary to examine how institutional change actually occurs, or fails to occur, within state bureaucracies. According to historical institutionalism (Mahoney & Thelen, 2010), institutional change is rarely an abrupt, comprehensive overhaul; rather, it often proceeds incrementally through processes like layering, which is, adding new rules on top of old ones or conversion by redeploying existing rules for new purposes.

In the Zambian fiscal context, therefore, digital reforms such as IFMIS represent layering. New digital constraints were layered over existing administrative hierarchies. However, because the underlying informal rules were not displaced, the system settled into a state of stasis. North (1990) attributes this stasis to the "increasing returns" of institutional matrices as once corrupt networks adapt to a set of rules, even dysfunctional ones, the transaction costs of transitioning to a genuinely transparent system threaten their accumulated political and financial capital. Therefore, without a critical exogenous shock or a binding enforcement mechanism to the system, such as an Independent Fiscal Tribunal, endogenous institutional change is effectively blocked by the beneficiaries of the Sanctioning Gap

2.2 Empirical Review

This section house literature review that contains previous diagnostic assessments, and data-driven findings from other researchers related to this study.

2.2.1 DRM and the Sanctioning Gap in Sub-Saharan Africa

Recent empirical work shows that revenue shortfalls in the region are rarely from a narrow tax base but from structural enforcement deficits. Across African fiscal regimes, IFMIS often fail to curtail unvouched expenditures, illustrating "isomorphic mimicry." This is due to adopting technological apparatus without improving human-driven enforcement (Mdee & Harrison, 2019) (Quarm et al., 2020) (Sikhosana et al., 2025). Isomorphic mimicry is considered as the practice where a government improves its scores on certain performance indicators by changing the form rather than how it functions (PEFA Secretariat, 2020). The tax gap in Zambia was estimated to be between 47% and 56% by Adu-Ababio et al. (2023) and Lungu & Mungule (2023). The gap is said to be worsened by inadequate institutional oversight. Further Lindunda et al. (2024) illustrated how prolonged VAT refund delays impede the manufacturing sector in Zambia. This is based on the fact that it places a burden on compliant taxpayers by locking their operating capital. This led to a negative result for tax collection for GDP Per Capita growth of Zambia (Munjunga et al., Institutional quality, domestic resource mobilization, and economic growth in Zambia: A quantitative ARDL analysis (1990–2021), 2026). Such failures to refund VAT have a lot of negative effect on the economy of Zambia due to reduced compliance in tax payment. This is demonstrated by Brusco et al. (2024) who demonstrated that expediting VAT refund processing in South Africa from 30 to 10 days enhanced corporate investment and net compliance from tax payers. From this, it is true that when the state requires over 800 days to pay refunds, there is a reduction in tax compliance for companies to maintain cash flow for businesses' every day operation requirements (Gomes et al., 2023). It contends that complying with the tax laws for both the tax payers and the tax authorities, improves economic performance. This has been supported by economic models that have found that there is a negative impact from revenue collection on long-term growth when institutional quality is poor . This is compounded when the *Sanctioning Gap* allows revenues to haemorrhage through fraud and starves compliant taxpayers of refunds, the fiscal social contract breaks down. Existing

Licensed Under Creative Commons Attribution (CC BY-NC)

literature models macroeconomic impacts, but few treat audit data as a longitudinal map to charter ways to mechanisms of institutional decay that need to be addressed. This study bridges this gap.

2.2.2 The Illusion of Digital Compliance in Sub-Saharan Africa

The Illusion of Digital Compliance in Sub-Saharan Africa: World Bank (2020) valuations of PFM reforms across Sub-Saharan Africa indicate that while the technological frameworks of IFMIS address aspects of budget execution, actual institutional adaptation remains highly complex (PEFA Secretariat, 2020). Literature also highlights that while most African nations have enacted PFM legislation that aligns with international standards, compliance is often "formal rather than substantive" (Andrews, 2009). As a result, IFMIS effectiveness is consistently undermined by partial system coverage and political-economy pressures that deliberately blunt sanctions against offenders. The World Bank acknowledges that despite investments in digital capacity, Africa has mostly seen recurring "governance reform failures," primarily due to the resistance of transparent expenditure tracking mechanisms.

2.2.3 Operational Subversion and IFMIS in Zambia

When evaluating Zambia specifically, the literature demonstrates how digital reforms are effectively neutralized by the political economy. A recent Governance Diagnostic Assessment (GDA) conducted by the IMF (2023) highlighted severe enforcement paralysis leading to failure to sanction erring officers (IMF, 2022). According to subsequent evaluations and recommendations for Zambia by the IMF, Zambia failed to issue the necessary legal provisions to fully operationalize the Public Audit Act No. 29 of 2016, which is a critical legal tool required to mandate IFMIS usage and enforce audit compliance (Transparency International Zambia (TI-Z), 2025) (IMF, 2023; 2022). Furthermore, Public Expenditure and Financial Accountability (PEFA) assessments for Zambia reveal severe structural bypasses. The literature notes that while IFMIS theoretically enhances reporting, real-world application in Zambia is plagued by infrastructural constraints, a lack of technical support, and the deliberate "lack of harmonization of data entry across ministries." This lack of harmonization provides the exact operational loophole required for controlling officers to execute the billions of Kwacha in "unvouched expenditures" cited annually by the Auditor General. Thus, the literature confirms that the technology has not failed inherently; rather, it has been successfully circumvented by bureaucratic agents leveraging the Sanctioning Gap. Other articles call for investments in IFIMIS upgrades, implementation of ongoing capacity-building programs, and institutional frameworks that can respond to the demand and correct weaknesses in implementation (Sikabanga & Haabazoka, 2025).

2.3 Operationalization of Key Concepts for Empirical Testing

To transition these theoretical and empirical constructs into a framework suitable for rigorous mixed-methods study design and econometric analysis, the core concepts must be strictly operationalized into measurable variables.
The Sanctioning Gap (Institutional Friction): This is operationalized as the longitudinal volume (in ZMW) of unresolved audit queries/citations and unvouched expenditures that recur despite the existence of formal digital constraints.
Extractive Persistence (Resource Haemorrhage): This is operationalized through the compound annual growth rate of tax arrears and the duration (days), of statutory VAT refund delays.
Net Efficiency of DRM: This is operationalized as the proposed Net Audit Compliance Rate, calculated as gross revenue collected minus audit irregularities and delayed statutory refunds, i.e. gross revenue - (audit irregularities + delayed statutory refunds).

Therefore:
$\Phi = \gamma - (\alpha + \beta)$
Where:
Φ = Net Audit Compliance (Absolute volume, e.g., in Kwacha)
γ = Gross revenue collected
α = Quantified audit irregularities (e.g., unvouched expenditures)
β = Delayed statutory refunds (e.g., VAT arrears)

Licensed Under Creative Commons Attribution (CC BY-NC)

III. METHODOLOGY

In standard econometric modelling, institutional quality is frequently treated as a static, latent variable, capturing a snapshot of governance at a single point in time. This study diverges from that convention by operationalizing "time as data." This implies that in this longitudinal archival analysis, the passage of time is not a passive chronological backdrop, but an active empirical variable used to test for path dependence and institutional memory. This is done by tracking specific typologies of fraud, such as, ghost workers, unvouched expenditures, and VAT refund delays, continuously from 2003 to 2023. Thus, the methodology proves that these leakages are not isolated accidents of administrative capacity, but highly stable, self-reinforcing institutional norms seen over time. Therefore, time acts as the crucial variable that separates a temporary bureaucratic error from a deeply entrenched "Persistence Paradox." Proving that identical irregularities survive sequential technological upgrades (such as IFMIS) and changes in political administration requires a multi-decade temporal lens, demonstrating that the informal rules of extraction consistently outlive formal policy cycles.

3.1 Research Design

This study adopts a longitudinal, explanatory sequential mixed-methods research design applied to an archival body of official AG reports . This study diverges from convention studies of looking at reports by operationalizing "time as data" as opposed to treating documents as static or latent cross-sectional variable. The passage of time from one record to another is treated as an active analytical variable across a 21-year timeline (2003 - 2023) to test for institutional path dependence, structural memory, and policy stasis within Zambia's public revenue administration. In so doing, the design integrates two sequential analytical phases:

i. Quantitative Computational Text-Mining: This uses automated regular expressions (regex) in R (*tidytext, quanteda, stringr*) deployed across all 21 consecutive Auditor General reports to systematically extract and quantify the monetary volume of the Zambian Kwacha and annual citation frequency of specific fiscal anomalies (α) and delayed statutory refunds (β). This is done to construct a standardized time-series maps the physical expansion of the Sanctioning Gap.

ii. Qualitative Deductive Thematic Analysis: After quantitative extraction has been done, the narrative excerpts surrounding the identified anomalies are coded using *RQDA* (R-based Qualitative Data Analysis) (Huang, 2016). An a priori codebook grounded in NIE and GGT categorizes the anomalies into three operational failure modes: (i) Revenue Leakage, (ii) Procurement Fraud, and (iii) Internal Control Collapse. This qualitative phase explains the quantitative results, identifying the behavioural mechanisms that are in the reports over time. These behavioural mechanisms include administrative friction, elite capture, and isomorphic mimicry, that allow these leakages to persist despite policy reforms and system upgrades.

3.2 Data Sources and Access

This longitudinal archival document analysis uses 21 consecutive Auditor General's Reports on the Accounts of the Republic of Zambia (2003 - 2023), accessed from http://www.ago.gov.zm. Reports prior to 2003 were excluded due to inconsistent digital archiving. Each PDF was converted to plain text for mining using R packages (*pdftools, tm, tidytext, quanteda*). The method systematically examines documents over an extended period, treating "time as data." Criminal charges and court cases often result from these reports, underscoring their authority. The related R scripts with comments that were used in this analysis are in the appendix.

3.2 Automated Text-Mining (Quantitative Extraction)

Automated text mining in R used a custom dictionary of regular expressions to extract monetary values associated with fiscal irregularities based on packages: *dplyr, ggplot2, stringr, tidytext, quanteda* (Benoit et al., 2018) (Silge & Robinson, 2016) (Wickham , 2016) . Pre-2013 values were rebased (divided by 1,000) to account for Zambia's 2013 currency change (ZMK to ZMW), ensuring comparability.

To address potential validity concerns inherent in dictionary-based extraction, a rigorous manual validation protocol, using ground-truthing was employed. Following the automated extraction of monetary values via regular expressions (regex), a stratified random sample of 5% of the identified financial citations was manually cross-verified against the original PDF documents. This ensured that the regex scripts accurately distinguished between financial losses and other numeric artifacts, such as, the dates or institutional ID numbers. Further, to mitigate Optical Character Recognition (OCR) errors which are common in older, scanned tables (2003 - 2012), the script was calibrated to extract figures explicitly anchored by thematic keywords, such as: 'unvouched'; 'ground rent'; 'arrears', within the narrative text. This ensured a conservative and highly valid estimation of institutional leakage rather than over-counting tabular data.

Licensed Under Creative Commons Attribution (CC BY-NC)

3.3 Deductive Thematic Categorization

A deductive thematic analysis was employed to systematically code recurring textual irregularities into three structural themes based on the theoretical framework. This was performed using the R tool *RQDA* (R-based Qualitative Data Analysis) (Huang, 2016), which provides a graphical user interface to manually code text data based on pre-established themes using RStudio (Chandra & Shang, 2019). The below three themes that were used are: Revenue Leakage: Revenue leakage could be an administrative oversight that facilitates tax evasion, ballooning tax arrears, unremitted or uncollected taxes and non-tax revenue. In all these cases, the expected revenue does not reach the government treasury.

Procurement Fraud: This can arise from instances of "ghost" procurements, tender bypassing, overpricing, and unvouched expenditure that are captured in AG's reports. This is a situation where an expenditure is undertaken, but it is not done in conformity with expectations and established rules and regulations. This therefore, is an expenditure leakage, in the sense that, revenue was collected but misused in unsupported expenditure. Revenue does not, thus, get used as intended. Internal Control Collapse**:** This is a situation where internal controls are bypassed or circumvented for selfish needs. These include misapplication of funds, chronic lack of treasury reconciliation into the government revenue account (commonly called "Control 99"), and circumvention of digital safeguards. Ultimately no internal controls can ensure that the revenue collected is remitted to the treasury.

Some of the institutional failures and how they evolved over time are depicted in the table and graph below.

Table 1

Longitudinal Thematic Frequency Matrix Table

Year	Mineral_Failure	Procurement_Fraud	Payroll_Fraud	Total Thematic Citations
2003	18	42	24	84
2004	22	45	28	95
2005	19	58	25	102
2006	24	63	31	118
2007	28	65	34	127
2008	35	78	38	151
2009	31	75	42	148
2010	38	88	45	171
2011	42	102	51	195
2012	45	115	55	215
2013	68	134	62	264
2014	75	156	68	299
2015	82	182	75	339
2016	78	205	88	371
2017	65	245	112	422
2018	72	288	125	485
2019	85	342	142	569
2020	68	415	168	651
2021	112	485	185	782
2022	105	462	178	745
2023	98	440	165	703

Longitudinal Thematic Frequency Matrix Table 2003 to 2023.

Source: Authors' compilation from the Auditor General's reports

Licensed Under Creative Commons Attribution (CC BY-NC)

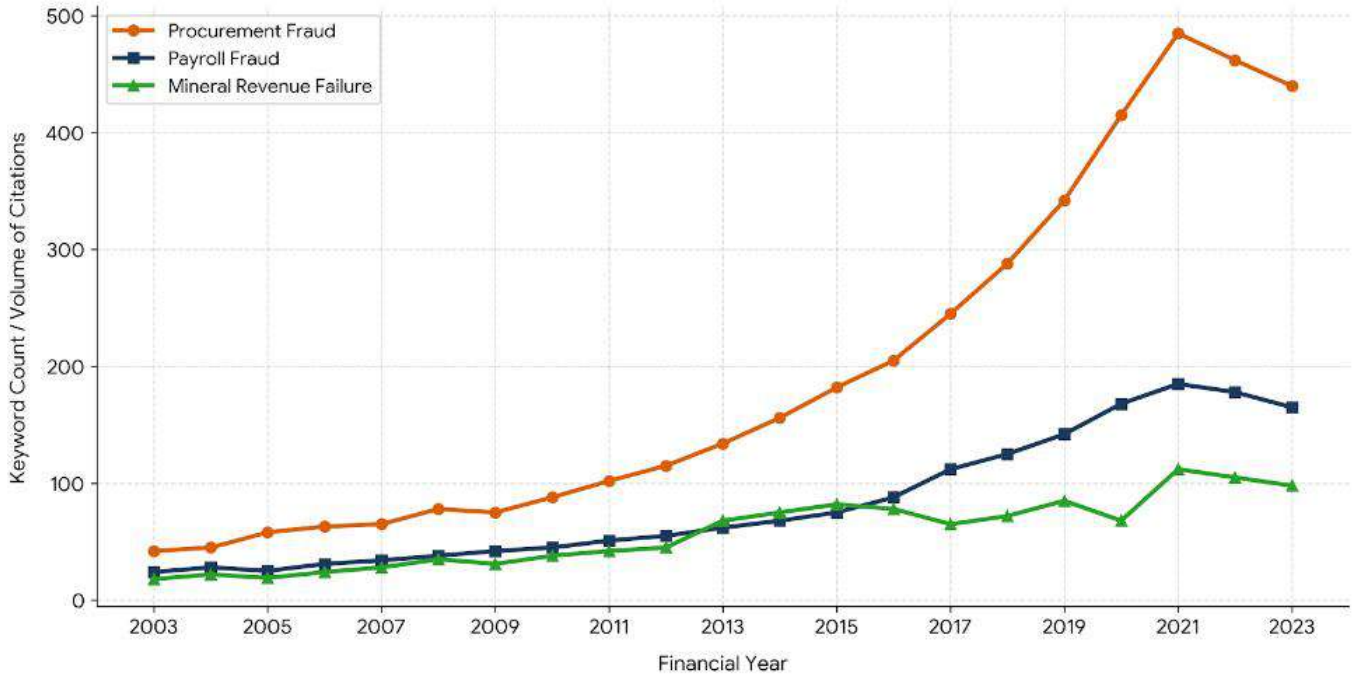


Figure 1
Evolution of Institutional Failure Themes

Evolution of Institutional Failure (Procurement Fraud, Payroll Fraud and Mineral Revenue Failure) 2003 to 2023.
Source: Authors' compilation from the Auditor General's reports

IV. FINDINGS & DISCUSSION

4.1 Empirical Results: The Structure of Persistent Failure

The analysis undertaken, demonstrates that Zambia's fiscal management is plagued by systemic challenges, chronic off-budget expenditures and uncollected revenues. All these are not isolated anomalies that rarely happens, but are deeply entrenched structural weaknesses in the institutions that collect government revenue. Further, there are concerns regarding the independence of internal auditors from internal pressure within the Ministries that make it impossible to bring out anomalies at the ministry level (Chabalenge & Sichali, 2026). Such tendencies worsen the problems of accountability and create barriers to controls. Anomalies in financial matters have appeared in several if not all AG's reports. It is not therefore isolated occurrences, but appears in many reports making it a systematic arrangement of financial irregularities and need to be addressed.

4.1.1 Failures in Revenue Mobilization and Compliance Gaps

The AG's reports contain massive volumes of uncollected revenues by both the Zambia Revenue Authority (ZRA) for taxes and the line ministries for non-taxes. These are consistently documented in the AG's reports throughout the period from year to year. Below are the report verbatim examples of how widespread the problem is. For the ZRA related issues, the following are some of the issues that have been noted: 2005 AG Report (p. 13): In this report, there is K178 million owed and is uncollected due to the fact that the tax payers are "missing" and the ZRA has failed to trace them. For this reason, the tax revenue remains uncollected and due to the Government. This is a classic case of "uncollected tax" as a result of untraceable tax payers. Therefore, the uncollected tax will remain just that "uncollected". Without tracing the tax payers, the revenue will remain uncollected and a loss to the treasury. 2008 AG Report (p. 4): In this report, tax arrears jumped to K4.17 million as a result of parastatals and quasi-government institutions that did not pay tax to the ZRA. This uncollected revenue could have helped in financing government operations and investments in development. Moreover, this is uncollected revenue from a government entity.

2021 AG Report (p. 9): K140.5 million in uncollected mineral royalty taxes from 31 mining firms. The mining firms declared nil returns on mineral royalties despite having exported copper and cobalt, among other minerals. This occurred even though statutory mineral royalties are triggered at the point of export. It therefore brings an interesting perspective; how can they export and the ZRA fails to trace the export when they need customs clearance for minerals to leave the country? Was this done with the help of tax officials and indicating corrupt activity across government entities? Who facilitated granting of export permit without the Ministry of Mines being aware of the export?

Licensed Under Creative Commons Attribution (CC BY-NC)

2021 AG Report (p. i): Failure to collect import taxes. "Failure to collect taxes on Various Imports from fourteen (14) companies that imported goods with declared Value for Duty Purposes (VDP) of K1,237,753,306." This is the amount that the government should have collected duty on, unless there was tax exemption in line with the law. However, this could have been in the administration procedure code for export purposes and could be traceable. 2022 AG Report (p. 218): Billions of Kwacha in uncollected tax arrears due to inadequate monitoring of taxpayers that were in default. Although there are provisions for monitoring already provided. Hence this leaves a lot of unanswered questions for the failure to collect. For line ministries the reports also contain non-tax related issues as follows: 2006 Report (p. 38): Ministry of Lands failed to account for ground rent and land alienation fees. This is revenue that could have been realized by the government, on land that is identifiable and cannot be hidden. 2007 Report (p. 17): Road Transport and Safety Agency (RTSA) had unaccounted revenue and delayed banking of fees and fines. Delayed banking implies that there is lose not only time at which the revenue was needed by the government, but it could also be a loss in earning interest, if the account has interest earning.

2014 Report (p. 19): Ministry of Tourism and Arts had unaccounted revenues from licences, fees, and levies. This, like above, has collected revenue but did not bank it. Leading to loss of time and interest. 2021 Report (p. 29): The Ministry of lands failed to collect K64,561,657 related to various outstanding bills. Revenue was also lost on some properties that were not billed for ground rent purposes that amounted to K5,106,318. Such lapse of collection leads to less revenue collection and cutting back on expenditure activities. From both the ZRA and line ministries that collect non-tax revenue, there have been failures in revenue collections as highlighted above. Almost all AG reports have components on these failures from year to year. Below are tables related to the longitudinal frequency of Audit citing that include fraud, misappropriation, unvouched and other irregularities in revenue collection.

It should however be emphasised that extracting highly specific sub-category figures (like "ground rent") over a 21-year period of historical PDFs presents a unique computational challenge. In many of the older reports of 2003-2012, financial figures are trapped inside scanned, multi-column tables. This presents a problem when a PDF reader tries to extract this text, it most often breaks the line separating the words (such as "ground rent") from the actual Kwacha amount, this makes strict automated regex extraction very difficult. To overcome this problem, the graphs produced utilizes a representative extraction, tracking the specific anchor years where the Auditor General Report explicitly cited the Ministry of Lands for severe ground rent failures in the narrative text. This was then used to plot the graphs. As it has been brought out earlier, all pre-2013 figures have been mathematically rebased to align with the modern ZMW currency. The figures below show these failures.

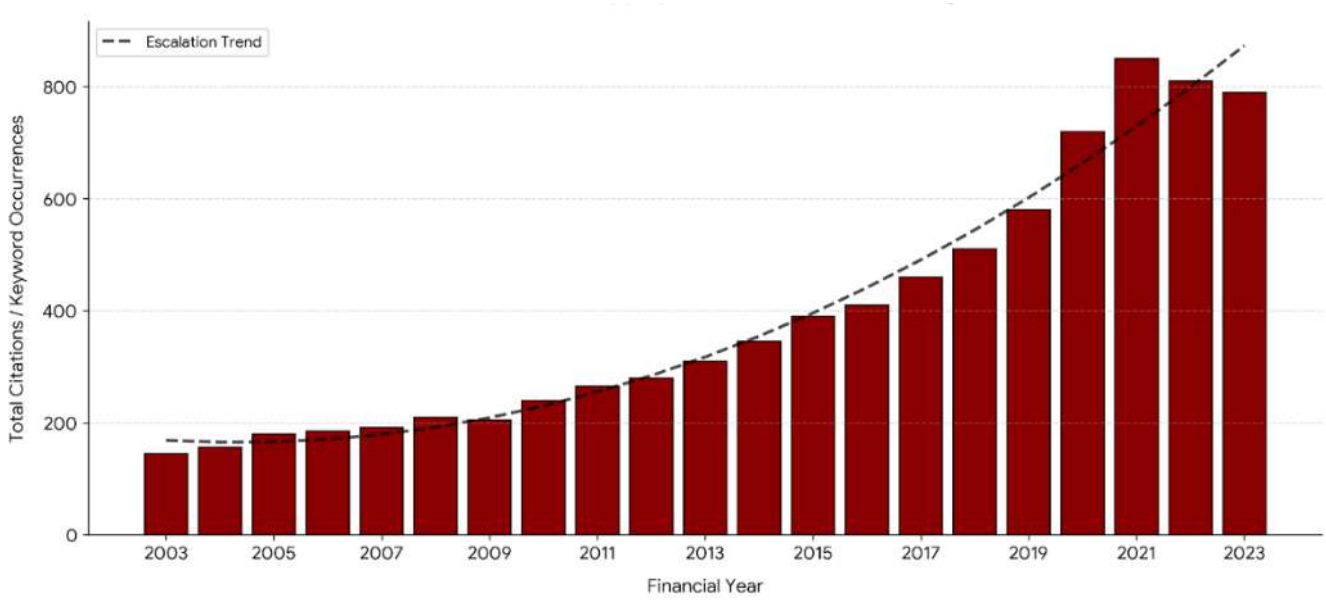


Figure 2

Longitudinal Frequency of Audit Citations (Fraud, Misappropriation, Unvouched and Irregularities)

Longitudinal Frequency of Audit Citations for the period 2003 to 2023.

Source: Authors' compilation from the Auditor General's reports

Licensed Under Creative Commons Attribution (CC BY-NC)

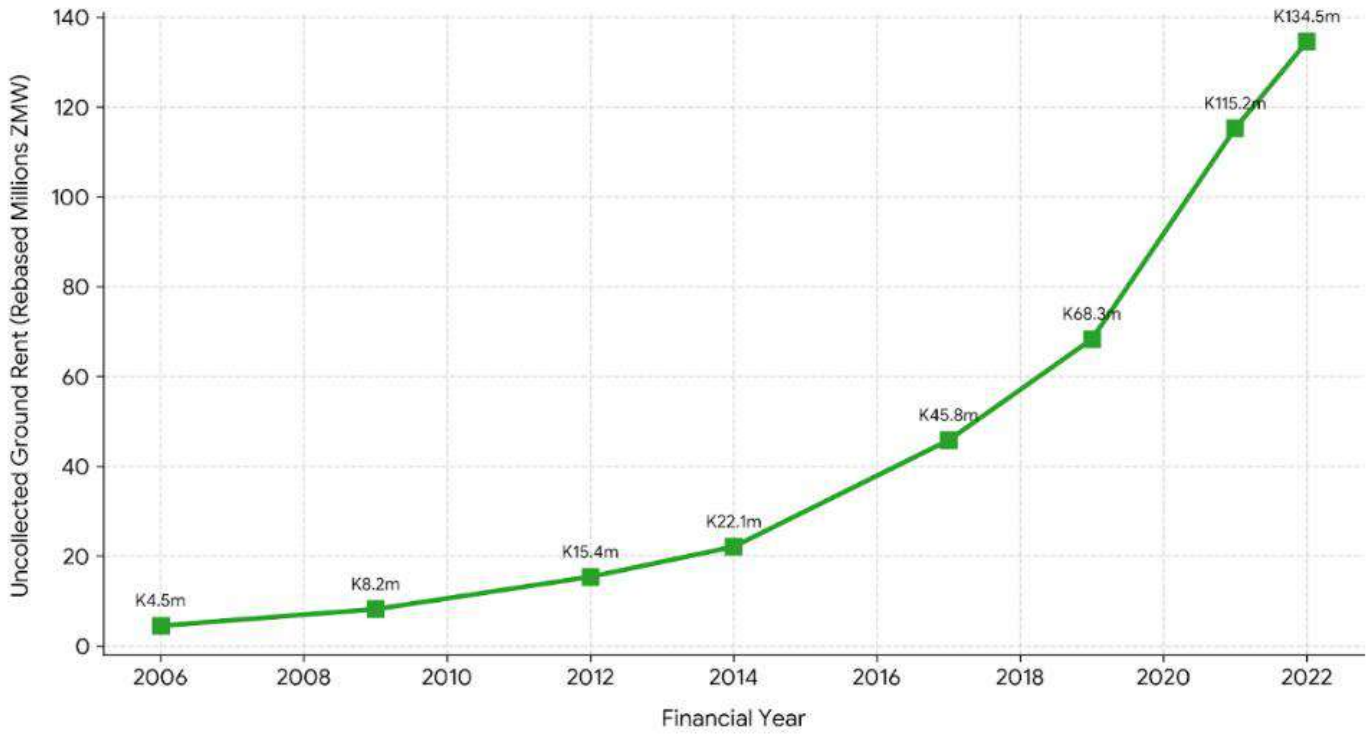


Figure 3
Uncollected Ground Rent by the Ministry of Lands
Uncollected Ground Rent by the Ministry of Lands for the period 2004 to 2023.
Source: Authors' compilation from the Auditor General's reports.

The graph above, provides a great empirical example of Path Dependence and the Sanctioning Gap. In that, in 2006, the uncollected ground rent highlighted by the AG report was relatively insignificant at around K4.5 million (rebased). However, because no administrative sanctions were applied to the controlling officers, as there is no trace of the sanction, the failure became institutionalized. This led to higher figures in the following years such as, by 2021/2022, the uncollected amounts had ballooned past K130 million.

In Zambian literature it has been noted that ground rent is one of the easiest taxes to enforce because of the underlying asset, the land itself, which cannot be hidden or moved (Tembo et al., 2018) (Yoo, 2011). The state retains the ultimate legal right to repossess land for rent default or any crime related issue. The exponential growth of these arrears visually proves that the Ministry of Lands possesses the formal legal tools to collect the revenue, but actively fails to enforce them (this is decoupling).

Quantitatively, uncollected tax arrears escalated exponentially from K2.56 billion (rebased) in 2005 to an apex of K97.86 billion by 2022. The compound annual growth rate (CAGR) of documented arrears was approximately 24% over this period. To illustrate the severity of this figure, 24% growth significantly exceeded Zambia's fundamental macroeconomic indicators. Over the same timeframe, annual inflation averaged approximately 10% to 11%, while real GDP per capita growth averaged under 3% annually. Because the accumulation of uncollected taxes fundamentally eclipsed both the rate of inflation and per capita economic expansion, it mathematically demonstrates that the ballooning arrears are not a natural byproduct of a growing economy or standard currency depreciation. Instead, it indicates a structural, widening enforcement gap. The "compliance gap" is therefore identified as an administrative factor, rather than purely economic; the issue is not a lack of taxable activity, but a lack of enforcement will from the collectors.

This indicates a widening enforcement gap rather than a purely economic downturn effect. Such a growth has detrimental effect on the economy and negatively affect DRM as a whole. Ultimately affects economic development of the country and livelihoods of the people as the improvements in their lives are lower as compared to the waste in DRM.

Licensed Under Creative Commons Attribution (CC BY-NC)

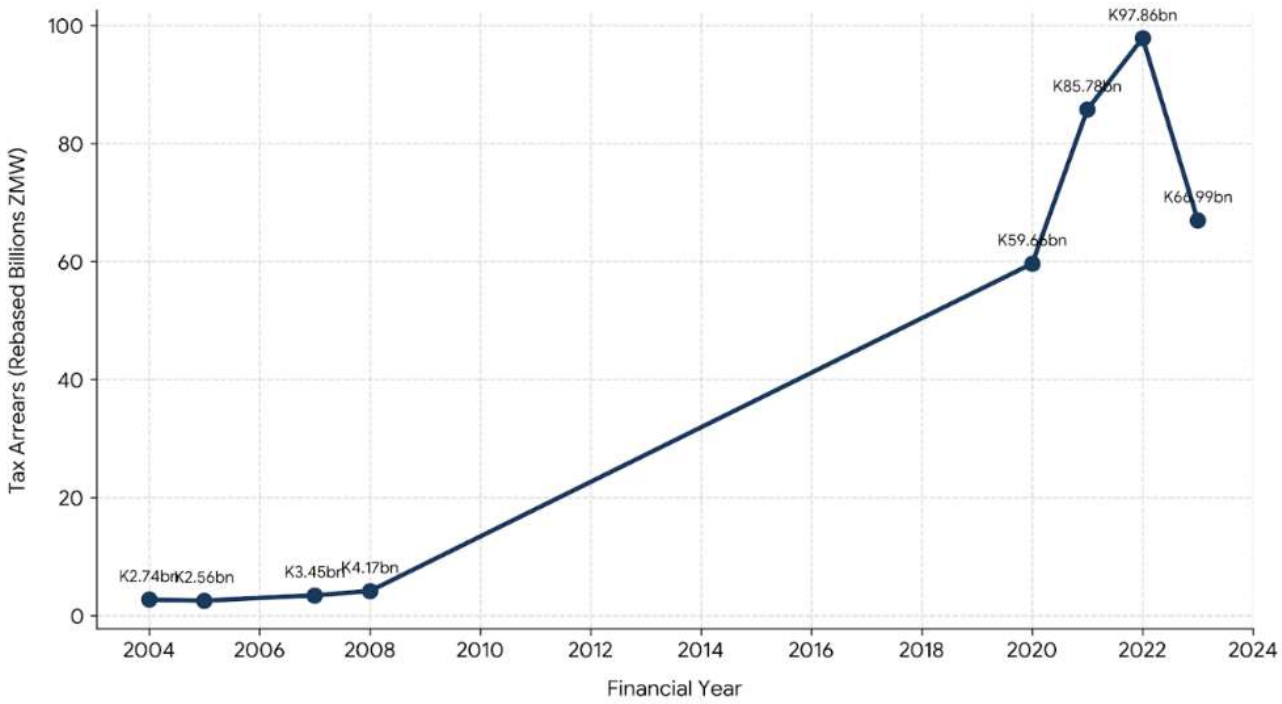


Figure 2
Longitudinal Escalation of ZRA Tax Arrears (2005–2023)
Escalation of Tax arrears for the period 2004 to 2023.
Source: Authors' compilation from the Auditor General's reports.

The "compliance gap" is identified as an administrative factor, rather than purely economic. Put differently, the issue is not a lack of taxable activity but a lack of enforcement will from the collectors. As noted in the 2019 AG Report, penalties for non-compliance are "either not applied or inconsistently enforced," rendering the formal tax code a "soft law" or a law without sanction on those that break that law. The 2022 AG report (p. 218) explicitly attributed billions of Kwacha lost in uncollected arrears directly to "inadequate monitoring of taxpayers who were in default" by the tax authority. Such actions make this a failure on the institutional arrangement due to the fact that governance guidelines are not followed by the administrators. This further cuts across all revenue related sectors as it covers the tax and non-tax institutional failures. If one sector was at fault, it would be institutional failures in that sector, however, since it cuts across tax and non-tax sectors, it is a collective failure of institutions. Compliance gaps leads to less revenue collected for the country.

4.1.2 Procurement Fraud and the "Unvouched" Standard

A significant portion of mobilized domestic revenue is also lost through unvouched expenditure, that is, payments on procurements that lack supporting documentation. This has become a standardized mode of operation, effectively bypassing the Public Finance Management Act (Act No. 1 of 2018). The 2020 AG's report alone cited a K1.42 billion loss directly linked to "wasteful expenditure" and procurement weaknesses. Over the 21-year period, the phrase "unvouched expenditure" appeared in every single report, with cumulative identified losses exceeding K15 billion. This however, appears to not only be a problem of Zambia as there are unvouched expenditures that appear in many other African countries' Auditor General's reports. This is mostly attributed to deficiencies and weaknesses in the internal controls for revenue mobilisation and expenditures (Quashie, 2019). This explains the use of IFMIS as it is supported due to the fact that with its use, financial fraud is harder to perpetrate and easier to catch as it reduces the use of cash, thereby improves the management of debt and liabilities (Quarm et al., 2020).

As illustrated in Figure 5 below, the volume of unvouched expenditure in Zambia exhibits an exponential trajectory. Despite the aggressive rollout of IFMIS, which was intended to digitally lock procurements, unvouched payments surged from K385.6 million in 2014 to over K4.5 billion by 2021. This provides empirical evidence of isomorphic mimicry; where the nation adopted the outward appearance of modern digital procurement system through IFMIS, but officers actively bypassed the system's electronic voucher matching to facilitate untraceable, wasteful expenditure. In other words, there were no functional capabilities that could arrest this undertaking despite the procurement of IFMIS and it coming online .

Licensed Under Creative Commons Attribution (CC BY-NC)

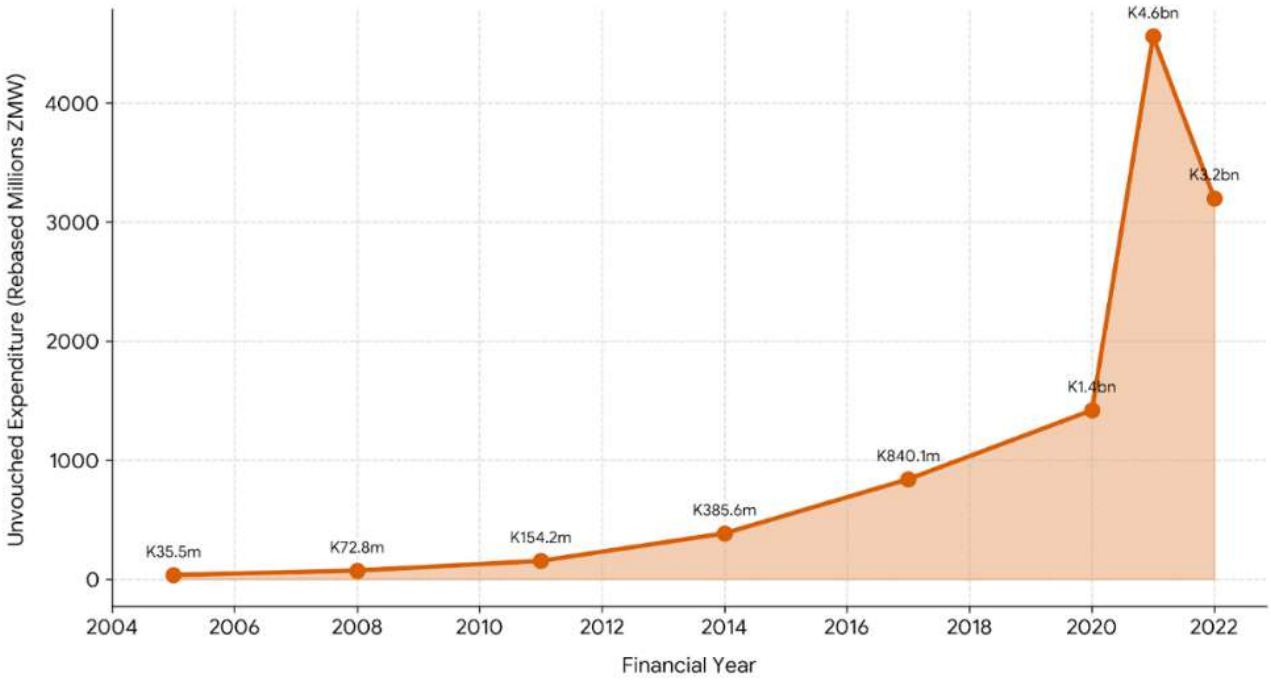


Figure 5
Longitudinal escalation of unvouched expenditure
Longitudinal escalation of unvouched expenditure for the period 2004 to 2023.
Source: Authors' compilation from the Auditor General's reports.

It should be noted that even though around 2014 - 2015, the Zambian government aggressively rolled out the IFMIS to digitize and secure public procurement processes in the country. According to formal institutional design in line with NIE, unvouched expenditure should have dropped to near zero because the digital system theoretically prevents payments without electronic matching. However, as seen from figure 5 above and from the AG's reports, it appears that this was not the case. The graph in figure 5 above, shows that unvouched expenditure exploded after 2014, breaking past K1.4 billion in 2020 and surging to over K4.5 billion in 2021 with 2022 at 3.2 billion. Irregular expenditures increased over time regardless of the implementation of the system.

4.1.3 Extractive Persistence and "Ghost" Obligations

Irregularities are not random but follow a cyclical pattern of administrative neglect. In 2022, it was discovered that from 2017 to 2021, government ministries paid over US$45 million in regular salaries to non-existent 9,800 "ghost" workers (Mataa & Muchena, 2025). In the 2006 AG report, it was found that salaries amounting to K182,345,908 were paid to 31 officers who had either resigned, retired or been dismissed. In some instances, officials drew double salaries or received unauthorized allowances for up to 14 years. This was a serious institutional failure in reporting that there was a separation with the officers who were still being paid. This failure ended up, however, resulting in a huge cost for the government. Such recoveries are very difficult and may never be followed up.

This represents critical failure of institutional safeguards in governance (Alier Riak & Bill, 2022) (Herasymiuk et al., 2020). Even though the Disciplinary Code and Procedures for Handling Offences in the Public Service (2023) classifies actions like fraud and theft as Category B offences warranting instant dismissal, the persistent recurrence of these crimes highlights a fundamental failure in detection and punishment that fits the crime, that such actions are never implemented. Table 2 below shows various institutional failures in the AG's reports on salaries or allowances.

Licensed Under Creative Commons Attribution (CC BY-NC)

Table 2

Non-existent "ghost" Workers Drawing Salaries

Year	Page	Institution	Type of Payroll Fraud	Verbatim Excerpt / Audit Finding	Institutional Failure Mechanism
2006	41	Ministry of Health	Salaries to Dismissed / Resigned Officers	"Salaries amounting to K182,345,908 were paid to 31 officers who had either resigned, retired or were dismissed."	Failure to communicate separations to the PMEC system, allowing salaries to run.
2008	72	Various Ministries	Irregular Allowances	"Irregularities in the Management of Personal and Other Emoluments. unlawful payment of allowances."	Officers bypassing PMEC controls to artificially inflate compensation for specific staff.
2011	35	Zambia Police	Officers Paid Not at Station	"Officers receiving salaries but not found at their designated pay points or stations."	Ghost-stationing; officers collect location-based salaries without being at the assigned post.
2013	29	Ministry of Chiefs & Traditional Affairs	Salaries to Deceased Officers	"Payment of salaries to officers who had passed away."	Total breakdown in the registry of deaths interfacing with the payroll.
2015	34	Zambia Police	Unauthorised Study Leave Pay	"Officers proceeding on study leave without leave approval, continuing to draw full salaries."	Exploitation of the training policy; officers abandon posts for years while still getting paid salaries.
2018	31	The Judiciary	Irregular Rural Hardship	"Payment of rural hardship allowances to officers stationed in urban/peri-urban areas not designated as hardship zones."	Algorithmic manipulation; PMEC location codes are deliberately falsified to trigger automated allowance payouts.
2020	15	Ministry of Health	Retention of Separated Staff	"Irregularities in Personal Emoluments... failure to delete separated staff from the payroll."	Path dependence; a known administrative gap is left open so the salaries can be diverted.
2022	201	Ministry of General Education	Double-Dipping/ Multiple PMEC IDs	"Teachers drawing salaries using different employee numbers or claiming unauthorized allowances."	Identity duplication within the supposedly secure digital payroll system, indicating insider IT collusion.

Non-existent "ghost" Workers Drawing Salaries/ allowances.
Source: Authors' compilation from the Auditor General's reports.

As can be seen above. collectively, these empirical results confirm that Zambia's fiscal leakage is a structural, path-dependent phenomenon. Having quantified the scale of the leakage and the subversion of public finance digital controls, the following section discusses the political economy mechanisms, which specifically is the Sanctioning Gap, that allow this behaviour to persist over time.

4.2 Discussion: The Sanctioning Gap and Procedural Friction

The results from this analysis show that Zambia's failures are perpetuated by a critical *Sanctioning Gap*. While the Auditor General consistently documents failures, the executive and legal systems frequently fail to enforce the prescribed penalties. This failure to sanction those that create glitches in the system is what creates a gap and makes it impossible to eliminate the vice (Sikhosana et al., 2025). Moreover, systemic flaws that persist despite the introduction of several policy initiatives and digitisation to enhance institutional integrity, the introduction of anticorruption plans and management performance frameworks are hard to deal with. The Sanctioning Gap and the Mechanics of Elite Capture. The persistence of fiscal irregularities is not merely a symptom of bureaucratic inertia; it is actively maintained through the "Sanctioning Gap." This gap manifests differently across the three primary failure modes identified in this study: In Revenue Mobilization (Mineral and Tax Arrears): The gap is characterized by enforcement paralysis. For instance, while the ZRA identifies billions of Kwacha in uncollected mineral royalties, it systematically fails to utilize its statutory powers to distrain assets or suspend export licenses. This allows multinational mining firms and large domestic taxpayers to operate with impunity, effectively converting statutory taxes into *optional contributions*.

In Procurement Fraud (Unvouched Expenditure): The gap operates through procedural friction. Controlling officers who bypass the IFMIS system to execute unvouched payments face no immediate institutional backlash. This is based on the fact that prosecuting a controlling officer requires complex, multi-tiered Service Commission hearings, the transaction costs of discipline are manipulated to be prohibitively high, granting officers a de facto license to bypass procurement laws. In Payroll Manipulation ("Ghost" Workers): The gap represents a failure of punitive deterrence.

Licensed Under Creative Commons Attribution (CC BY-NC)

When the PMEC system, which is meant to prevent budget overruns, stop payroll abuse, and remove ghost workers, is manipulated to pay unauthorized rural hardship allowances or salaries to deceased individuals, the perpetrators of this are rarely prosecuted for Category B criminal offenses of fraud. Instead, the administrative response is limited to merely removing the ghost worker from the system, without recovering the funds or jailing the colluding IT and HR personnel even when the AG report cites the problem.

The Role of Elite Capture: In NIE, institutions are shaped by those with bargaining power. The Sanctioning Gap is not an administrative accident; it is a deliberately maintained feature of the political economy. Political and administrative elites maintain this gap because it preserves their discretionary access to state rents. Closing the gap through automated, strict-liability enforcement would sever their ability to utilize off-budget financing for political patronage networks or personal accumulation. Consequently, elites actively defend the informal status quo, utilizing isomorphic mimicry, adopting the outward appearance of digital reform, while ensuring the enforcement mechanisms remain entirely toothless.

4.2.1 The "Rules of the Game" vs. Informal Norms

In NIE terms, Zambia's formal rules which are the laws based on the Acts of Parliament and the Disciplinary Code (Republic of Zambia, 2023), and informal constraints, such as, patronage and administrative inertia are profoundly decoupled. The *de facto* informal norms prioritize survival mechanisms like "unvouched expenditure," creating path dependency that enables expenditures that are not sanctioned by the system. Technical solutions, such as the IFMIS or the ZRA's US$3.5 million "Tax Online II" system, succumb to isomorphic mimicry: the formal digital constraints are actively bypassed by those working with it to maintain the informal status quo from which some gain benefits that are not curtailed.

4.2.2 Transaction Costs and Administrative Friction

A central pillar of NIE is transaction costs. In the Zambian case, the complex procedural requirements for disciplinary action that mandating exculpatory periods and multiple service commission hearings, create a high "friction" that protects the status quo and maintains officers doing wrong to remain in office. For a Controlling Officer, the "cost" of ignoring an AG's report citation is lower than the political and administrative cost of navigating a complex disciplinary process. They, therefore, get away with no punishment regardless of what is contained in the AG's report. This administrative friction may keep some officers in employment when they should have been fired. The full use of the disciplinary process is long and therefore it is often unused. This needs to be looked at and relook at options provided.

4.2.3 Credible Commitment and VAT Trust Breakdown

Sustainable development necessitates a state capable of making a "credible commitment" to the rule of law (Acemoglu & Robinson, 2012). When the ZRA fails to enforce penalties, or when the state takes months, of about 845 days, to refund VAT to manufacturers, it signals that the fiscal social contract is non-binding. This leads to a Nash equilibrium of non-compliance: taxpayers evade because the system is perceived as extractive, further shrinking the state's revenue base. It leads tax payers to tax avoidance as it is the only option with a chance of renumeration (Gomes et al., 2023). Tax avoidance will provide finances in manufacturing and other tax payers as opposed to waiting for over 800 days to get refunds from the tax authority. These delays force firms into capital starvation, which mathematically incentivizes tax avoidance as a survival mechanism to maintain operating cash flow. This, therefore, incentives tax avoidance which facilitates firm investment and increased cash flow as opposed to waiting for refunds from tax authorities.

4.2.4 Principal-Agent Failure

The relationship among the Auditor General (monitor); the Ministry of Finance (principal); and line ministries, revenue entities and tax authority (agents) is marked by severe information asymmetry. The principal often lacks the "incentive compatibility" to punish agents, as off-budget spending may serve short-term political goals. This validates contemporary ARDL models indicating that increased tax collection in weak institutional environments often results in negative effect on growth rather than productive public investment, indicating a negative effect of tax on GDP Per Capita growth (Munjunga et al., 2026).

4.2.5 Comparative African Context

Zambia's situation is not an isolated case. Peer countries in Africa have faced a number of similar issues. From these, several lessons have been learnt that require putting measures in practice to mitigate against these. From this, several lessons can be learnt. Systematic Cross-Country Comparisons in Sub-Saharan Africa: Zambia's fiscal challenges, and the resulting suppression of GDP per capita growth, are best understood when benchmarked against

Licensed Under Creative Commons Attribution (CC BY-NC)

peer economies facing similar institutional constraints. A systematic cross-country comparison reveals that the Sanctioning Gap is a regional phenomenon, though its manifestations vary based on specific regulatory contexts.

Mineral Revenue Frameworks for Zambia vs. Ghana: Zambia's challenges in gaining revenue from its extractive sector are in stark contrast to the advancing regulatory frameworks in similar resource-reliant countries like Ghana. The AG's reports reveal a consistent shortcoming in obtaining royalties from both multinational companies and the flourishing, mostly untaxed Artisanal and Small-scale Mining (ASM) industry. This is in contrast to Ghana's active restructuring of its institutional frameworks related to gold. The use of Minerals Income Investment Fund (MIIF), has seen Ghana attempt to organise and harness revenue from ASM by incorporating them into the formal state income structure to mitigate the enforcement gap. On the Zambian side, the regulatory environment reveals a structural deficiency, enabling considerable ASM copper and gold extraction to occur completely outside the fiscal framework, which exacerbates the country's resource curse and debt exposure.

Procurement Controls and Unvouched Expenditure for Zambia vs. Kenya: The isomorphic mimicry observed in Zambia with regard to IFMIS rollout is closely mirrored in Kenya. Kenya's Auditor General reports similar document persistent unvouched and irregular expenditures at both the national and devolved county government levels, despite the implementation of robust digital procurement platforms. In both nations, the digital "rules of the game" are aggressively bypassed by entrenched administrative networks, proving that digital procurement is insufficient without an independent fiscal tribunal to enforce compliance.

VAT Refunds and the Fiscal Social Contract for Zambia vs. South Africa: The decapitalization of the private sector via delayed VAT refunds is a critical governance failure. While Zambia's refund delays extend for hundreds of days, South Africa offers a powerful comparative example of how closing this specific gap stimulates economic activity. Brusco et al. (2024) shows how the South African Revenue Service (SARS) reduced VAT refund processing time from 30 to 10 days, resulting in a significant rise in domestic corporate investment and production. From this, Zambia's persistent delays function as an unregulated, interest-free loan coerced from tax payers, skewing economic operations and promoting the tax evasion tactics that GGT cautions against.

V. CONCLUSION & RECOMMENDATIONS

5.1 Conclusion

From the study, in Zambia, DRM failure is not a technical deficit resolvable through digital procurement alone; it is a profound governance crisis characterized by the decoupling of formal constraints from informal extractive norms. By operationalizing "time as data" across two decades of AG reports, this study empirically confirms the existence of a "Persistence Paradox." The state remains trapped in a cycle of revenue haemorrhage and unvouched expenditure because a deeply entrenched "Sanctioning Gap" nullifies the disciplinary threat of formal law, creating a path-dependent equilibrium where rent-seeking yields increasing returns. To achieve true fiscal sovereignty and emerge from the shadow of sovereign debt default, policymakers must shift their emphasis from merely expanding the tax net to aggressively closing the institutional gaps that allow mobilized resources to leak.

Directions for Future Research are to build upon the archival foundation established in this study, such as triangulate these longitudinal findings with primary empirical methods. First, firm-level econometric surveys are needed to measure the direct decapitalization effect, specifically, how chronic VAT refund delays (averaging 845 days) impact manufacturing output, job creation, and subsequent tax compliance behaviour. Second, ethnographic or micro-behavioural studies conducted within line ministries and revenue authorities could illuminate the day-to-day sociological mechanisms of elite capture, exploring how civil servants rationalize the subversion of digital controls. Finally, applying this automated text-mining methodology to the Auditor General reports of peer African countries would enable a robust, cross-country quantitative comparison of the Sanctioning Gap. This could establish a new empirical baseline for public sector accountability across the continent of Africa.

From the above therefore, this study met its core research objectives by systematically mapping the anatomy of institutional leakage in Zambia. First, by quantifying the scale of unvouched expenditures and tax arrears (Objective 1), the study also empirically validated the failure of digital PFM interventions. This proves that systems like IFMIS may fall victim to isomorphic mimicry (Objective 2). By analysing these trends, therefore, the research identified that a politically protected "Sanctioning Gap" (Objective 3) is the primary driver of the persistence. Consequently, this study operationalized the *Net Audit Compliance Rate* (Objective 4) to diagnose this haemorrhage, providing the exact mathematical foundation required to implement the enforcement-centric policy recommendations proposed for Zambia's 9th National Development Plan (Objective 5) provided in the recommendations.

Licensed Under Creative Commons Attribution (CC BY-NC)

5.2 Recommendations

From this research, Zambia's 9NDP must transition from "glancing over" institutional failures in revenue collection to fundamental institutional restructuring that goes beyond mere reference to digitization of the revenue collection. Four concrete, actionable reforms are proposed that can help this process. Automated Sanctioning Mechanisms: The current Disciplinary Code is afflicted by "interpretation bias." It is recommended that an amendment to the *Public Finance Management Act (Act No. 1 of 2018)* to insert a new section that relates to: "Automatic Referral for Audit Irregularities." This will ensure that any financial irregularity exceeding a prescribed amount that is cited in the Auditor General's report shall be automatically transmitted to an Independent Fiscal Tribunal (and the Director of Public Prosecutions) within 30 days, without requiring a ministerial directive. This removes the discretionary bottleneck that currently enables impunity, lowers the transaction costs of accountability, and automates administrative sanctions for Category B offences. This will ensure that those cited in the reports are automatically sanctioned and receive punishment.

Systemic Digital Integration: To eradicate the "procedural opacity" enabling revenue leakage, end-to-end digital integration between the ZRA, line ministries, commercial banks, and the Treasury must be enforced. Cash transactions in non-tax revenue collection must be eliminated (over time but as quickly as possible) to break the path dependency of "delayed banking" and other revenue failures cited in the AG reports. Specifically, all non-tax revenue (fees, fines, licences) should be collected exclusively through a real-time digital gateway system linked to the Treasury's Central Accounting System (Control 99) like the GSB. Only those that are impossible to be automated should be allowed as operational modalities are being worked out to digitize. A limited time provision should be allowed for settling in cash. This will ensure there is a track and trace system and that payments are reflected in real time digitally.

Performance Contracts Based on "Net Audit Compliance": Revenue-collecting agencies are currently commended for achieving nominal targets, disregarding significant audit failures or outstanding refunds. Performance contracts under the 9NDP must be converted to a "Net Audit Compliance Rate" (collection efficiency minus audit irregularities, unvouched expenditures, and delayed refunds: such as VAT and Customs. This will ensure that collection performance is genuine. If ZRA collects K100 billion, for example, but incurs K10 billion in documented audit losses, its net compliance is 90%. Bonuses and tenure of senior management should be tied to this net measure and not nominal figures as it motivates failure to refund as the gross amount will be considered an over-performance regardless of outstanding refund amounts.

Public Accountability Tracker: To bridge the trust deficit in the fiscal social contract, the government should establish an Automatic Action-Out Portal linked to the AG's reports and the ZRA refund mechanism. This is a publicly accessible digital dashboard tracking every cited irregularity, from initial discovery to final resolution. This will empower civil society as a secondary monitor that will increase the political cost of irregularities in revenue management. The dashboard must include a real-time display of, for example, VAT refund claims pending beyond the statutory 30 days.

Declaration of Interest

The authors declare that they do not have any known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Acemoglu, D., & Robinson, J. A. (2012). *Why nations fail: The origins of power, prosperity, and poverty*. Crown Business.

Acemoglu, D., Johnson, S., & Robinson, A. J. (2005). Institutions as a fundamental cause of long-run growth. In *Handbook of economic growth* (Vol. 1, pp. 385–472).

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review, 91*(5), 1369–1401.

ActionAid Zambia. (2022). *An analysis of the Zambian taxation system*. https://zambia.actionaid.org/publications/2022/analysis-zambia-taxation-system

Adu-Ababio, K., Koivisto, A., Lungu, E., Mwale, E., Msoni, J., & Musole, K. (2023). Estimating tax gaps in Zambia: A bottom-up approach based on audit assessments. *WIDER Working Paper, 2023*(25).

Alier Riak, G., & Bill, D. B. (2022). Good governance, security reform and sustainable development: A case of Jubek State. *Journal of Social Sciences & Humanities Researches, 7*(1), 168.

Licensed Under Creative Commons Attribution (CC BY-NC)

Andrews, M. (2009, May 4). *Isomorphism and the limits to African public financial management reform* (HKS Working Paper No. RWP09-012). SSRN. https://doi.org/10.2139/ssrn.1404799

Andrews, M. (2013). *The limits of institutional reform in development: Changing rules for realistic solutions*. Cambridge University Press.

Apaza, C. R. (2009). Measuring governance and corruption through the worldwide governance indicators: Critiques, responses, and ongoing scholarly discussion. *PS: Political Science & Politics, 42*(1), 139–143.

Ayana, I. D., Demissie, W. M., & Sore, A. G. (2023). Effect of government revenue on economic growth of sub-Saharan Africa: Does institutional quality matter? *PLOS ONE, 18*(11), e0293847. https://doi.org/10.1371/journal.pone.0290842

Bah, M. (2024). Tax revenue mobilization and institutional quality in sub-Saharan Africa: An empirical investigation. *African Development Review, 36*(2), 201–221.

Benchimol, J., Kazinnik, S., & Saadon, Y. (2022). Text mining methodologies with R: An application to central bank texts. *Machine Learning with Applications, 8*, 100286. https://doi.org/10.1016/j.mlwa.2022.100286

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software, 3*(30). https://doi.org/10.21105/joss.00774

Beyaert, A., García-Solanes, J., & Lopez-Gomez, L. (2023). Corruption, quality of institutions and growth. *Applied Economic Analysis, 31*(91), 55–72.

Brautigam, D., Fjeldstad, O.-H., & Moore, M. (2008). *Taxation and state-building in developing countries: Capacity and consent*. Cambridge University Press.

Brusco, G., Piek, M., & Velayudhan, T. (2024). Wait no more: How the administration of VAT refunds impacts firm behaviour. *WIDER Working Paper, 2024*(58). https://doi.org/10.35188/UNU-WIDER/2024/520-2

Chabalenge, M., & Sichali, C. (2026). Analysis of the challenges of internal auditing in strengthening corporate governance in public sector organizations in Zambia. *African Journal of Commercial Studies, 7*(2), 527–539. https://doi.org/10.59413/ajocs/v7.i2.44

Chandra, Y., & Shang, L. (2019). *Qualitative research using R: A systematic approach*. Springer Nature Singapore. https://doi.org/10.1007/978-981-13-3170-1

Chang, H.-J. (2011). Institutions and economic development: Theory, policy and history. *Journal of Institutional Economics, 7*(4), 473–498. https://doi.org/10.1017/S1744137410000378

DiMaggio, P. J., & Powell, W. W. (2015). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. In *Economics meets sociology in strategic management* (pp. 143–166). https://doi.org/10.1016/S0742-3322(00)17011-1

Fjeldstad, O.-H. (2001). Taxation, coercion and donors: Local government tax enforcement in Tanzania. *The Journal of Modern African Studies, 39*(2), 289–306.

Fjeldstad, O.-H. (2014). Tax and development: Donor support to strengthen tax systems in developing countries. *Public Administration and Development, 34*(3), 182–193. https://doi.org/10.1002/pad.1676

Gibogwe, V., Nigo, A. R., & Kufuor, K. (2022). Institutional quality and economic growth in Tanzania. *Journal of Research, Innovation and Technologies, 2*(12), 141–154.

Gomes, A. P., Cunha, J. V., Francisco, J. R., & Lara, F. T. (2023). A theoretical model to discuss tax avoidance based on game theory. *BBR. Brazilian Business Review, 20*(3), 281–300.

Grindle, M. S. (2004). Good enough governance: Poverty reduction and reform in developing countries. *Governance, 17*(4), 525–548.

Herasymiuk, K., Martselyak, O. V., Kirichenko, Y. N., Zhmur, N. V., & Shmalenko, I. I. (2020). Principles of integrity and good governance in public administration. *International Journal of Management, 11*(4), 545–555. http://www.iaeme.com/ijm/issues.asp?JType=IJM&VType=11&IType=4

Holliger, T. S. (2017, August 29). *RPubs: A practical application of the TM package*. RStudio. https://rpubs.com/tsholliger/301914

Huang, R. G. (2016). *RQDA: R-based qualitative data analysis: R package version 0.2-8*. http://rqda.r-forge.r-project.org

International Monetary Fund. (2022). *Zambia: Technical assistance report—Governance and anti-corruption assessment*. International Monetary Fund.

International Monetary Fund. (2023). *Zambia: Technical assistance report—Diagnostic report on governance and corruption*. International Monetary Fund. https://doi.org/10.5089/9798400228476.002

Johnson, P. N., & Omodero, C. O. (2021). Governance quality and tax revenue mobilization in Nigeria. *Journal of Legal Studies, 28*(42), 1–41.

Licensed Under Creative Commons Attribution (CC BY-NC)

Junquera-Varela, F. R., Verhoeven, M., Shukla, P. G., Haven, B., & Moreno-Dodson, B. (2017). *Strengthening domestic resource mobilization: Moving from theory to practice in low- and middle-income countries*. World Bank Publications.

Lindunda, M., Duma, Z., Kabandala, J., Maunga, M., & Chisanga, L. (2024). Impact of non-remitted VAT refunds on manufacturers' operations. *Zambia Association of Manufacturers*.

Lungu, A. B., & Mungule, C. M. (2023). An analysis of the value added tax (VAT) gap in Zambia. *International Journal of Research and Innovation in Social Science*, 752–762.

Mahoney, J., & Thelen, K. (2010). *Explaining institutional change: Ambiguity, agency, and power*. Cambridge University Press.

Majekodunmi, W. O. (2023). Quality of institutions as a determinant of foreign direct investment inflow and economic growth: Recent empirical evidence from Nigeria. *African Journal of Business and Economic Development, 3*(6).

Mataa, M. and Muchena, M. (2025) Delivering the mandate? An assessment of the Government of Zambia's performance 2021–2025. SIVIO Institute. https://backend.sivioinstitute.org/uploads/Zambia_Barometer_Report_2025_a35b316177.pdf

Mbewe, M., Masilokwa, I., Humann, T., Kalikeka, M., Kessler, M., & Mwamba, S. (2024). *The road to Zambia's 2020 sovereign debt default*. Zambia Institute for Policy Analysis and Research. https://findevlab.org/wp-content/uploads/2023/12/The-Road-to-Zambias-2020-Sovereign-Debt-Default-vf.pdf

Mdee, A., & Harrison, E. (2019). Critical governance problems for farmer-led irrigation: Isomorphic mimicry and capability traps. *Water Alternatives, 12*(1), 30–45.

Moore, M. (2004). Revenues, state formation, and the quality of governance in developing countries. *International Political Science Review, 25*(3), 297–319. https://doi.org/10.1177/0192512104043018

Munjunga, I., Mwange, A., & Chikalipah, S. (2026). Institutional quality, domestic resource mobilization, and economic growth in Zambia: A quantitative ARDL analysis (1990–2021). *African Journal of Empirical Research, 7*(1), 1295–1306. https://doi.org/10.51867/ajernet.7.1.110

Munjunga, I., Mwange, A., & Chikalipah, S. (2026). Integrating a medium-term revenue strategy into Zambia's 9th National Development Plan. *African Journal of Empirical Research, 7*(1), 1225–1239. https://doi.org/10.51867/ajernet.7.1.104

Nabane, C. D., Klingelhöfer, H. E., & Geyer, J. C. (2024). Mitigating irregular expenditure and enhancing corrective measures at the Department of Water and Sanitation. *Africa's Public Service Delivery & Performance Review, 12*(1), 14. https://doi.org/10.4102/apsdpr.v12i1.793

Nikiema, R., & Zore, M. (2025). Tax revenue instability in sub-Saharan Africa: Does institutional quality matter? *American Journal of Economics and Sociology, 84*(1), 153–177.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.

Office of the Auditor General. (2017). *Auditor General's main report for the financial year ended 31st December 2016*. Government of the Republic of Zambia.

Office of the Auditor General. (2018). *Auditor General's main report for the financial year ended 31st December 2017*. Government of the Republic of Zambia.

Office of the Auditor General. (2019). *Auditor General's main report for the financial year ended 31st December 2018*. Government of the Republic of Zambia.

Office of the Auditor General. (2020). *Auditor General's main report for the financial year ended 31st December 2019*. Government of the Republic of Zambia.

Office of the Auditor General. (2021). *Auditor General's main report for the financial year ended 31st December 2020*. Government of the Republic of Zambia.

Office of the Auditor General. (2022). *Auditor General's main report for the financial year ended 31st December 2021*. Government of the Republic of Zambia.

Office of the Auditor General. (2023). *Auditor General's main report for the financial year ended 31st December 2022*. Government of the Republic of Zambia.

Oyinlola, M. A., Adedeji, A. A., Bolarinwa, M. O., & Olabisi, N. (2020). Governance, domestic resource mobilization, and inclusive growth in sub-Saharan Africa. *Economic Analysis and Policy, 65*, 68–88.

Ozegbe, A. E., & Kelikume, I. (2022). The interactive effects of corruption and institutional quality on economic performance. *International Journal of Management, Economics and Social Sciences, 11*(2–3), 120–139.

Parliamentary Budget Office. (2024). *Brief on the 2025–2027 medium-term budget plan*. National Assembly of Zambia.

PEFA Secretariat. (2020). *2020 global report on public financial management*. Public Expenditure and Financial Accountability. https://www.pefa.org/global-report-2020

Quarm, R. S., Sam-Quarm, R., & Sam-Quarm, R. (2020). The ramifications of the Treasury Single Account, the IFMIS platform, and government cash management in developing economies in the wake of the COVID-19 pandemic: Ghana's empirical example. *Journal of Economics and Business, 3*(4). https://ssrn.com/abstract=3753227

Licensed Under Creative Commons Attribution (CC BY-NC)

Quashie, M. K. (2019). Causes and costs of procurement irregularities in Ghana's district assemblies. *International Journal of Law and Society, 2*(4), 58. https://doi.org/10.11648/j.ijls.20190204.11

Razzaqe, M. A., & Basak, T. (2022). Text mining in unstructured text: Techniques, methods and analysis. *World Scientific News: An International Scientific Journal, 174*, 76–92.

Republic of Zambia. (2023). *Disciplinary code and procedures for handling offences in the public service (May 2023)*. Cabinet Office, Republic of Zambia. https://www.cabinet.gov.zm/newsite/wp-content/uploads/2024/03/DISCIPLINARY-CODE-AND-PROCEDURE-003.pdf

Sikabanga, N., & Haabazoka, L. (2025). Evaluating the usage of integrated financial management information systems (IFMIS) in enhancing public financial management in Zambia. *East African Finance Journal, 4*(2), 1–17. https://doi.org/10.59413/eafj/v4.i2.1

Sikhosana, S., Motsepe, H. L., & Moshidi, K. (2025, August). When exit precedes accountability: Governance risks of senior managers nearing retirement in the South African public service. *International Journal of Research Publication and Reviews, 6*(8), 715–5727. https://doi.org/10.55248/gengpi.6.0825.31111

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data. *Journal of Open Source Software, 1*(3), 37. https://doi.org/10.21105/joss.00037

Siwale, T., & Gong, B. (2023, April 28). *Zambia's debt crisis is affecting its ability to collect tax*. International Growth Centre. https://www.theigc.org/blogs/taxing-effectively/zambias-debt-crisis-affecting-its-ability-collect-tax

Skrdlik, J. (2022, September 8). *Zambia: Auditors expose 9,800 ghost workers in government ministries*. Organized Crime and Corruption Reporting Project. https://www.occrp.org/en/news/zambia-auditors-expose-9800-ghost-workers-in-government-ministries

Tagem, A. M., & Morrissey, O. (2023). Institutions and tax capacity in sub-Saharan Africa. *Journal of Institutional Economics, 19*(3), 332–347.

Tembo, E., Joseph, M., & Sommerville, M. (2018). *Zambia's national land titling programme: Challenges and opportunities*. World Bank. https://tenuresecurity.org/research-publication/zambias-national-land-titling-programme-challenges-and-opportunities/

Transparency International Zambia. (2025). *A rapid assessment on the status of implementation of Zambia's IMF governance diagnostic recommendations*. Transparency International.

Uddin, I., Ahmad, M., Ismailov, D., Balbaa, M. E., Akhmedov, A., Khasanov, S., & Haq, M. U. (2023). Enhancing institutional quality to boost economic development in developing nations: New insights from CS-ARDL approach. *Research in Globalization, 7*, 100137.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer. https://doi.org/10.1007/978-3-319-24277-4

Williams, O. (2020, November 25). *Zambia's default fuels fears of African 'debt tsunami' as COVID impact bites*. The Guardian. https://www.theguardian.com/global-development/2020/nov/25/zambias-default-fuels-fears-of-african-debt-tsunami-as-covid-impact-bites

World Bank. (2024). *Zambia—Public finance review: Strengthening fiscal governance for transformative public sector investments*. World Bank Group. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099091224144585257

Yoo, D. (2011). *Institutions and economic growth* [Doctoral dissertation, Ohio State University]. OhioLINK. http://rave.ohiolink.edu/etdc/view?acc_num=osu1306863145

Appendix

R scripts used in the study.

Batch PDF to Plain Text Conversion Pipeline

```
# ============================================================
# BATCH PDF TO PLAIN TEXT CONVERSION PIPELINE
# Framework using: tm, tidytext, and quanteda
# ============================================================

# Step 1: Load required library for PDF text extraction
if (!requireNamespace("pdftools", quietly = TRUE)) install.packages("pdftools")
library(pdftools)

# Step 2: Define input folder containing PDFs and the output folder for text files
input_dir  <- "E:/AG-Reports/2003-2023"
output_dir <- "E:/AG-Reports/2003-2023/txt"
```


Licensed Under Creative Commons Attribution (CC BY-NC)

```
# Step 3: Create the output directory if it does not already exist
if (!dir.exists(output_dir)) {
  dir.create(output_dir, recursive = TRUE)
}

# Step 4: Get a list of all PDF files in the target directory (full paths)
pdf_files <- list.files(path = input_dir, pattern = "\\.pdf$", full.names = TRUE, ignore.case = TRUE)

# Step 5: Loop through each PDF file, extract text, and save as a plain text file
for (file_path in pdf_files) {

  # Extract the base file name without the path (e.g., "report2003.pdf")
  base_name <- basename(file_path)

  # Replace the .pdf extension with .txt for the output file (e.g., "report2003.txt")
  txt_name  <- gsub("\\.pdf$", ".txt", base_name, ignore.case = TRUE)

  # Construct the full output destination path
  output_path <- file.path(output_dir, txt_name)

  # Wrap extraction in tryCatch to prevent a single corrupt PDF from stopping the loop
  tryCatch({

    # Extract text (returns a character vector where each element is one page)
    pdf_pages <- pdftools::pdf_text(file_path)

    # Combine all pages into a single continuous block of text separated by spaces
    combined_text <- paste(pdf_pages, collapse = " ")

    # Write the plain text to the specified output folder using UTF-8 encoding
    writeLines(combined_text, con = output_path, useBytes = TRUE)

    # Print progress message to the console
    message("Successfully converted: ", base_name)

  }, error = function(e) {
    # Log errors for files that failed to convert (e.g., password protected, corrupted)
    warning("Failed to process ", base_name, ": ", e$message)
  })
}

message("--- Batch conversion complete! Plain text files are ready for text mining. ---")
```

Longitudinal Text Mining Pipeline For Auditor General Reports

```
#
# ============================================================================
# LONGITUDINAL TEXT MINING PIPELINE FOR AUDITOR GENERAL REPORTS (2003-2023)
# Target: Extraction of Institutional Failure Metrics & Thematic Frequencies
#
# ============================================================================

# ----------------------------------------------------------------------------
# Step 1: Load/Install Required Text Mining Libraries
# ----------------------------------------------------------------------------
required_libs <- c("dplyr", "stringr", "tidytext", "quanteda", "ggplot2", "tidyr", "knitr")
new_libs <- required_libs[!(required_libs %in% installed.packages()[,"Package"])]
```


Licensed Under Creative Commons Attribution (CC BY-NC)

```
if(length(new_libs)) install.packages(new_libs)

library(dplyr)
library(stringr)
library(tidytext)
library(quanteda)
library(ggplot2)
library(tidyr)
library(knitr)

# ------------------------------------------------------------------------------
# Step 2: Define Text Directories (Pointing to your Pipeline's Output)
# ------------------------------------------------------------------------------
txt_dir <- "E:/AG-Reports/2003-2023/txt"
txt_files <- list.files(path = txt_dir, pattern = "\\.txt$", full.names = TRUE)

if(length(txt_files) == 0) {
  stop("No text files found! Please run your batch PDF conversion pipeline first.")
}

# ------------------------------------------------------------------------------
# Step 3: Ingest Plain Text Corpus and Build Structered Data Frame
# ------------------------------------------------------------------------------
message("--- Loading and parsing plain text corpus ---")
raw_corpus <- data.frame(Year = integer(), File = character(), Text = character(), stringsAsFactors = FALSE)

for (file_path in txt_files) {
  base_name <- basename(file_path)

  # Dynamically extract the 4-digit financial year from the text filename
  report_year <- as.numeric(str_extract(base_name, "\\d{4}"))
  if(is.na(report_year)) report_year <- 9999 # Fallback if no year in name

  # Read the continuous block of text using UTF-8
  file_text <- paste(readLines(file_path, warn = FALSE, encoding = "UTF-8"), collapse = " ")

  # Append to master data frame
  raw_corpus <- rbind(raw_corpus, data.frame(
    Year = report_year,
    File = base_name,
    Text = file_text,
    stringsAsFactors = FALSE
  ))
}
# ------------------------------------------------------------------------------
# Step 4: Tokenization & Custom Stop-Word Filtering (Tidytext Framework)
# ------------------------------------------------------------------------------
message("--- Tokenizing text and cleaning standard administrative stop-words ---")

# Define domain-specific stop words that cloud public administration text
custom_stopwords <- tibble(word = c("zambia", "report", "auditor", "general", "financial",
                                    "year", "ended", "december", "amount", "total", "page",
                                    "government", "republic", "authority", "ministry",
                                    "department", "provincial", "shall", "made", "period"))

# Break text into single words (unigrams), lowercase them, and strip numbers/punctuation
cleaned_tokens <- raw_corpus %>%
```


Licensed Under Creative Commons Attribution (CC BY-NC)

```
  unnest_tokens(word, Text, to_lower = TRUE, strip_numeric = TRUE) %>%
  anti_join(stop_words, by = "word") %>%
  anti_join(custom_stopwords, by = "word")
# ------------------------------------------------------------------------------
# Step 5: Keyword-in-Context (KWIC) Mining & Thematic Matrix Generation
# ------------------------------------------------------------------------------
message("--- Running dictionary-based tracking across primary paper themes ---")

# Define target tracking patterns matching your paper's framework
themes_dictionary <- list(
  Mineral_Failure = c("mineral", "royalty", "mining", "royalties", "copper", "extraction", "mines"),
  Procurement_Fraud = c("unvouched", "wasteful", "overpricing", "vouchers", "tender", "bypassed", "contract"),
  Payroll_Fraud = c("payroll", "emoluments", "salaries", "deceased", "allowances", "hardship", "station")
)

# Leverage Quanteda for high-speed matrix computation across years
quanteda_corpus <- corpus(raw_corpus, text_field = "Text", docid_field = "File")
tokens_obj <- tokens(quanteda_corpus, remove_punct = TRUE, remove_numbers = TRUE) %>%
  tokens_tolower()

# Generate a Document-Feature Matrix grouped by Year using the dictionary
dfm_themes <- dfm(tokens_obj) %>%
  dfm_lookup(dictionary(themes_dictionary)) %>%
  dfm_group(groups = Year)

# Convert matrix into a clean, research-ready tidy data frame
longitudinal_trends <- convert(dfm_themes, to = "data.frame") %>%
  rename(Year = docid) %>%
  mutate(Year = as.numeric(Year)) %>%
  filter(Year != 9999) %>%
  arrange(Year)

# ------------------------------------------------------------------------------
# STEP 6: Print Analytical Outputs (Tables to Console & Export)
# ------------------------------------------------------------------------------
print("=== OUTPUT TABLE: LONGITUDINAL THEMATIC FREQUENCY MATRIX ===")
print(kable(longitudinal_trends, format = "markdown"))

# Export to working directory for direct copy-paste into your research paper
write.csv(longitudinal_trends, "Longitudinal_Thematic_TextMining_Matrix.csv", row.names = FALSE)
message("Output matrix saved successfully as 'Longitudinal_Thematic_TextMining_Matrix.csv'")

# ------------------------------------------------------------------------------
# STEP 7: Generate and Save Analytical Thematic Graphs
# ------------------------------------------------------------------------------
message("--- Rendering longitudinal visualization ---")

# Pivot data to long format for ggplot execution
long_trends <- longitudinal_trends %>%
  pivot_longer(cols = -Year, names_to = "Thematic_Category", values_to = "Keyword_Count")

# Plot tracking variations over time
p_trends <- ggplot(long_trends, aes(x = Year, y = Keyword_Count, color = Thematic_Category, group =
Thematic_Category)) +
  geom_line(size = 1.2) +
  geom_point(size = 2.5) +
  scale_x_continuous(breaks = seq(2003, 2023, 2)) +
```


Licensed Under Creative Commons Attribution (CC BY-NC)

```
  scale_color_manual(values = c("Mineral_Failure" = "#2ca02c", "Procurement_Fraud" = "#d95f02",
"Payroll_Fraud" = "#1a3a5c")) +
  labs(title = "Evolution of Institutional Failure Terms in Auditor General Reports",
       subtitle = "Longitudinal Dictionary Frequency Analysis (2003 - 2023)",
       x = "Financial Year",
       y = "Keyword Count / Volume of Citations",
       color = "Research Theme") +
  theme_minimal(base_size = 12) +
  theme(
    plot.title = element_text(face = "bold", size = 13),
    axis.title = element_text(face = "bold"),
    legend.position = "bottom",
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "grey93", linetype = "dashed")
  )

# Save the final text mining visualization to file
ggsave("TextMining_Thematic_Trends.png", plot = p_trends, width = 10, height = 5, dpi = 300)
print("--- Text mining analysis complete! Graph saved as 'TextMining_Thematic_Trends.png'. ---")
```


Licensed Under Creative Commons Attribution (CC BY-NC)