



South Africa's Tax Gaps

Where is revenue being lost?

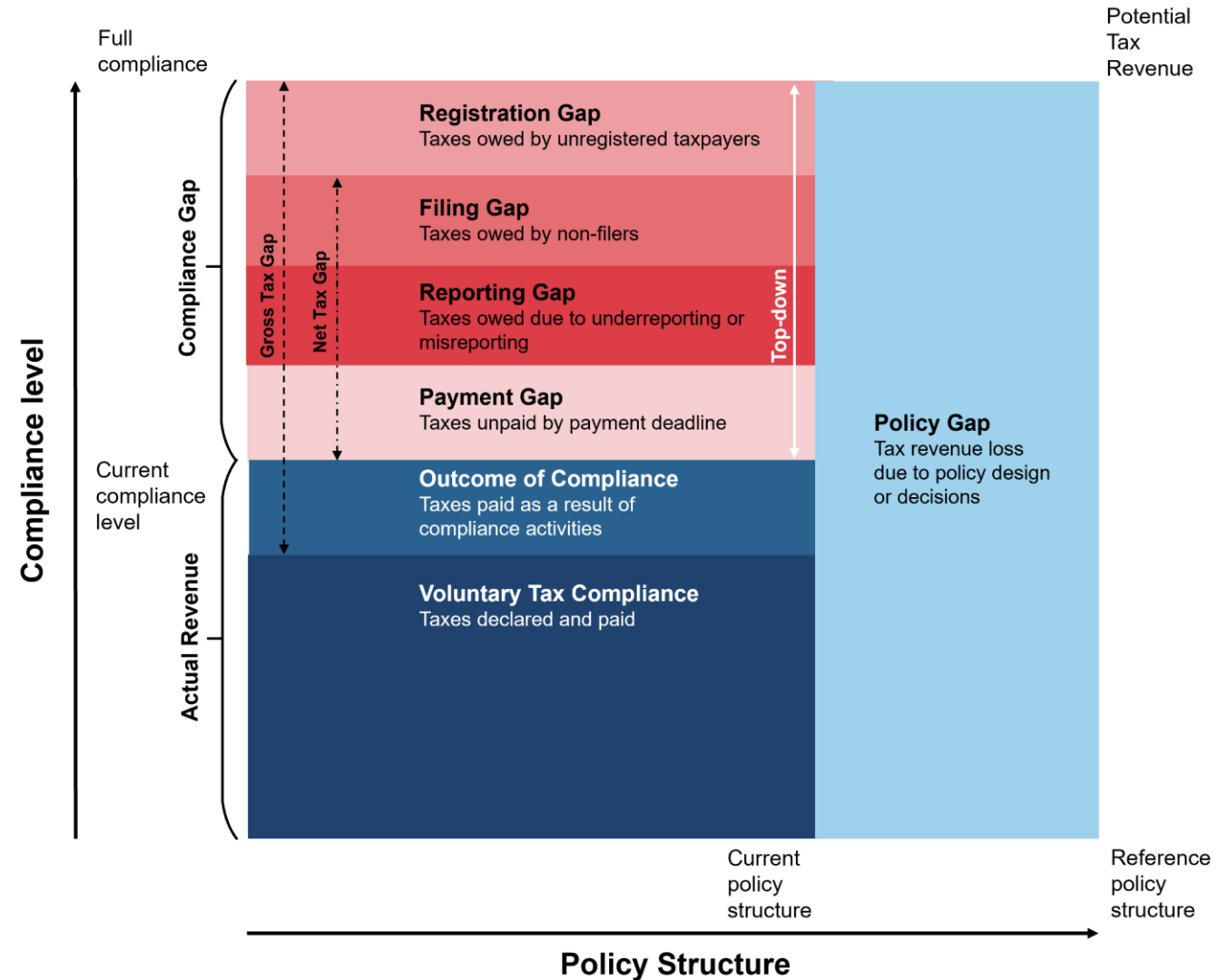
Amina Ebrahim | Research Fellow, UNU-WIDER
GTAC weekly webinar | 9 Oct 2026

SA-TIED

Southern Africa – Towards Inclusive Economic Development

Defining the Tax Gap

- **Policy Gap**
- **Compliance Gap**
 - Payment gap
 - Reporting gap
 - Filing Gap
 - Registration Gap
- Two main methodologies
 1. Top down (macro, aggregate)
 2. Bottom up (micro, detailed)



Where can additional revenue come from?

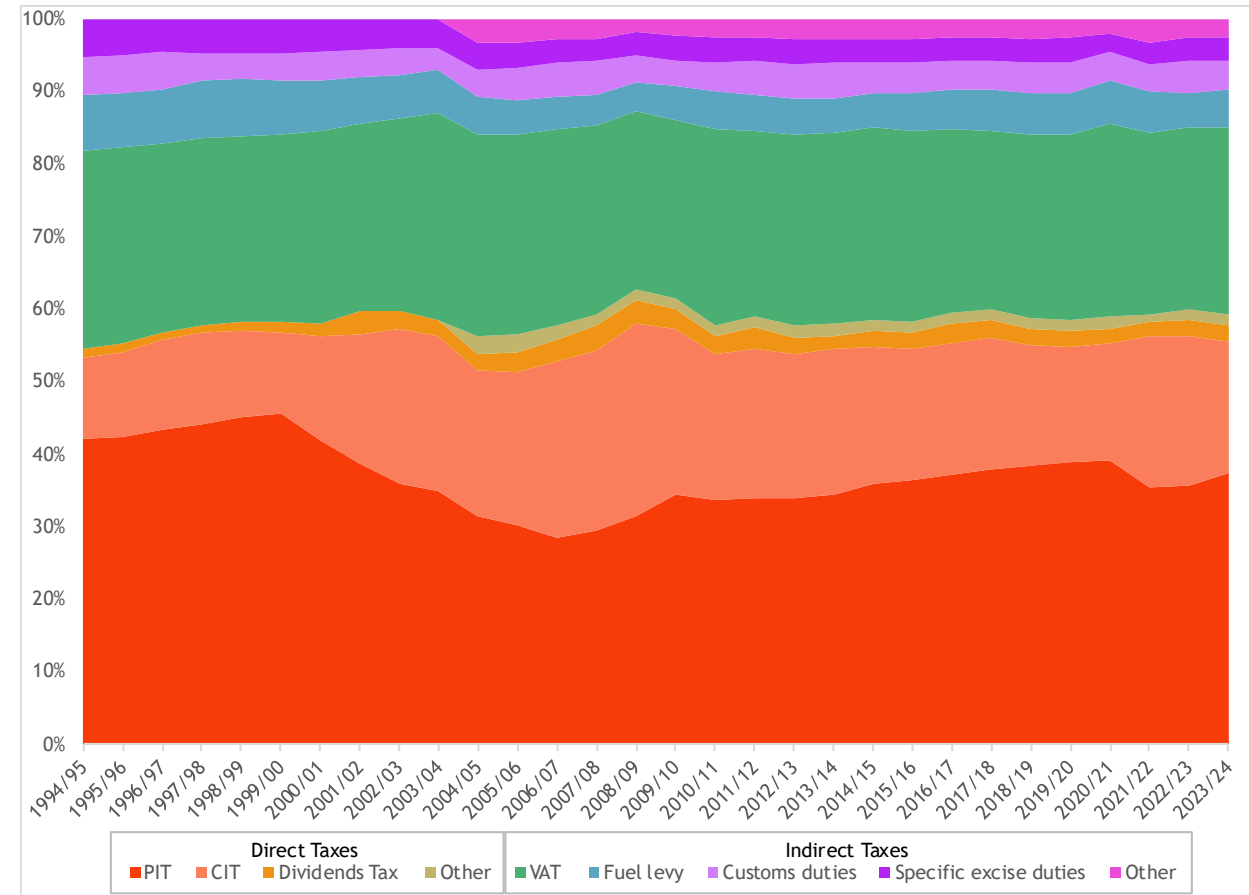
- Growth expands the tax base.
- Tax policy changes legal liabilities, thereby affecting the **Policy Gap**.
- Better compliance collects more of what the current law already requires, thereby affecting the **Compliance Gap**.
- Tax Gap estimates help locate the channels and show where policy and administration design matters.

How do we do this?

- VAT and CIT returns tell us what firms declared. Audits reveal some discrepancies between declarations and assessed liability.
- We use those observed cases to estimate underreporting among other filers. Registration, non-filing, payment and policy gaps remain outside these estimates.

Background

- Corporate income tax is SA's 3rd largest revenue source, contributing roughly 17% of total tax revenue and about 4.3% of GDP.
- VAT contributes ~25% to tax revenue and our VAT-to-GDP ratio is 6%–7%
- South Africa has lacked systematic, firm-level evidence on where non-compliance is concentrated and how it varies across different types of businesses.



Methodology – Bottom-up approach

1. Observe declarations across firms and tax years.
2. Learn from audit outcomes, accounting for differences among taxpayers.
3. Predict the reporting gap for firms and periods without an audit.
4. Aggregate across the population of **reporting** taxpayers.

Machine Learning - Random Forest algorithm

Two stages:

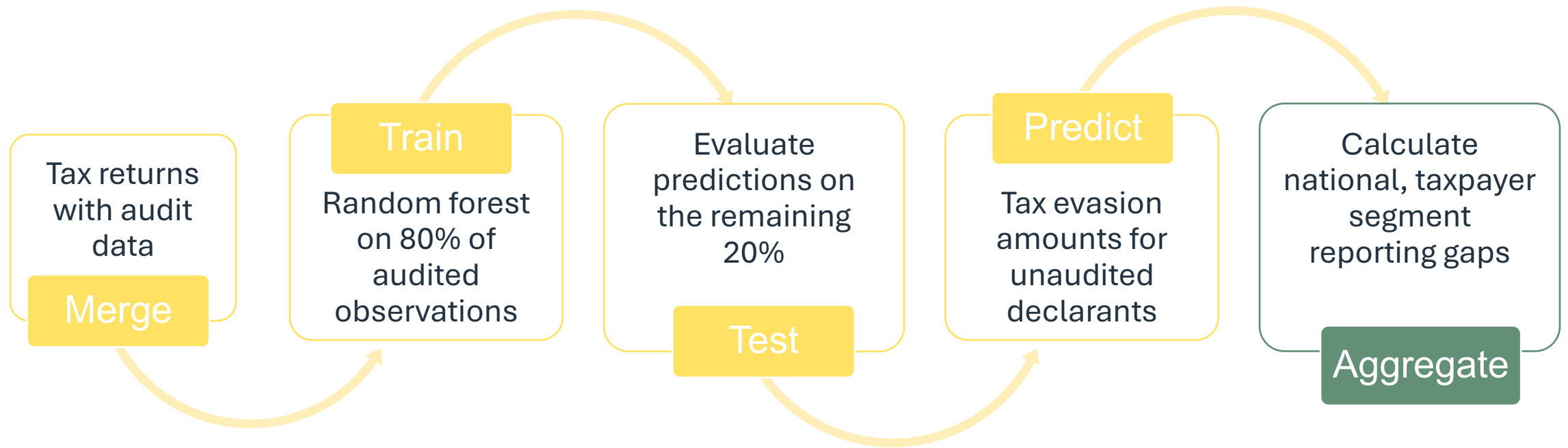
1. **predict** the probability of compliant vs non-compliant
2. if non-compliant - **predict the adjustment amount**

$$Evasion_{it} = f(Tax\ Return\ Variables_{it}, Firm\ Characteristics_{it})$$

$$potential\ tax = Tax\ **reported** + predicted\ tax\ evasion$$

$$Total\ Tax\ Gap = \frac{\sum_{pop} tax\ evasion}{\sum_{pop} potential\ tax} \times 100$$

More about the methodology



ML Tax Gap in other countries

- Italy: Applied a 2-step Gradient Boosting approach.
- **Armenia**: Estimated CIT using XGBoost, and PIT and SSC gaps (IMF).
- Rwanda: new ML paper in CESifo Journal
- Zambia: Applied Artificial Neural Network approach to estimate both CIT and VAT
- Tanzania: Applied Random Forest
- France: Applied Gradient Boosting (*in French*)
- Sweden: Applied Random Forest

Data

- Tax Returns: annual CIT and periodic VAT returns
- Audit Data: Specialised Audits (Issue audits and Comprehensive Audits). Exclude verifications.
- Model includes predictors from the return data:
 - Key firm characteristics
 - Key financial measures (including ratios)
 - To capture dynamic trends, we include both the **deviation from the industry** average and the **year-on-year change** for key variables.

Focusing on the **compliance gap**

CIT Reporting Gap

13.5% average estimated CIT reporting gap, 2016–2020.

25% **Small firm tax gap** | Average across firms with revenue below R20m

~2% **Medium/large firm tax gap** | Average across larger incorporated businesses

>20% **Construction sector share** | Of the estimated aggregate corporate tax gap

VAT Reporting Gap

40.6% average estimated VAT reporting gap, 2016–2020.

Construction, wholesale & retail are among the sectors showing the largest VAT gap.

>50% gaps seen in *Mpumalanga* and the *Northern Cape*.

CIT Compliance reporting gap

By year and firm size

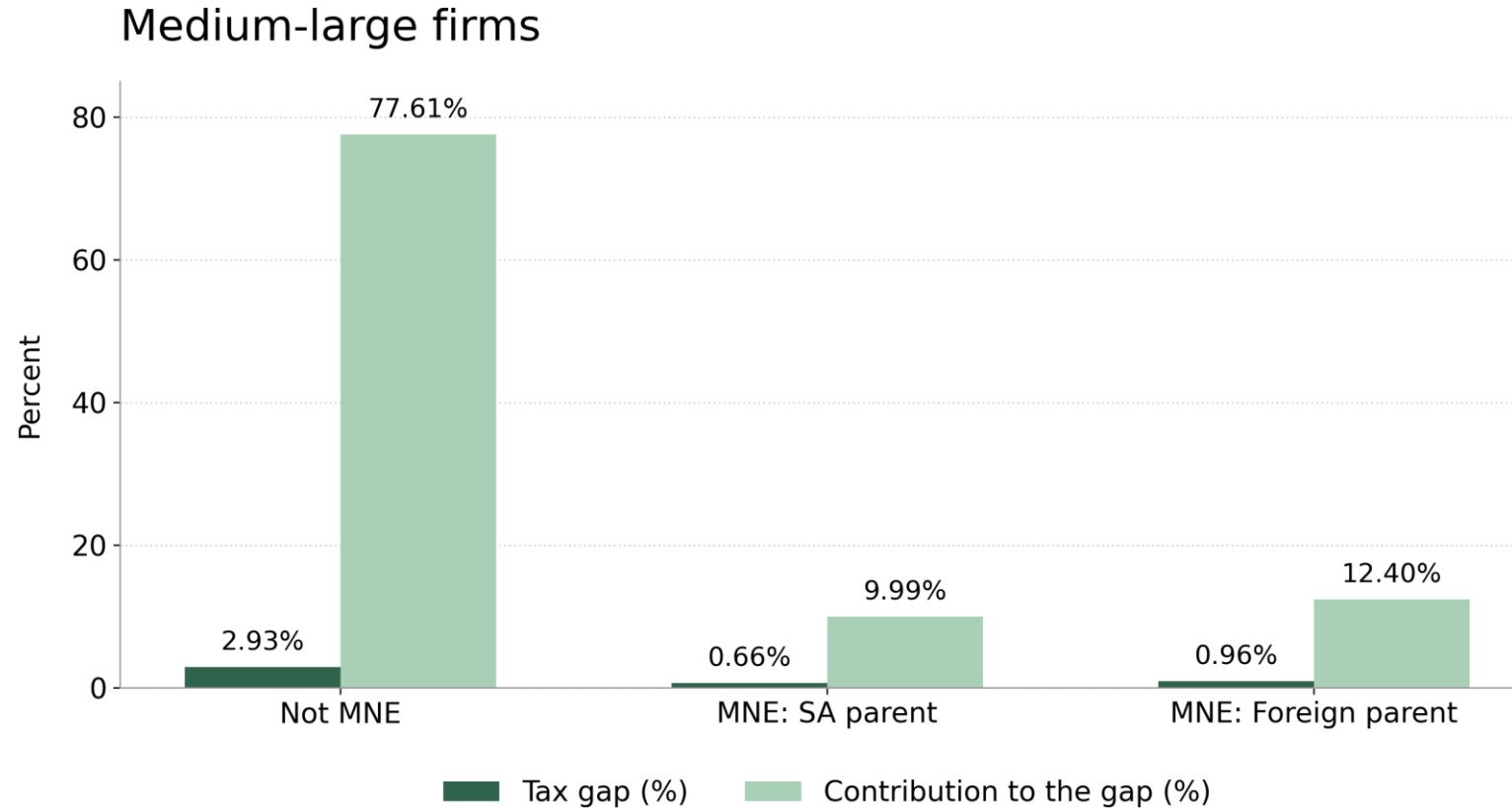


SA-TIED

Southern Africa – Towards Inclusive Economic Development

CIT Compliance reporting gap

Medium-large firms, by MNE status



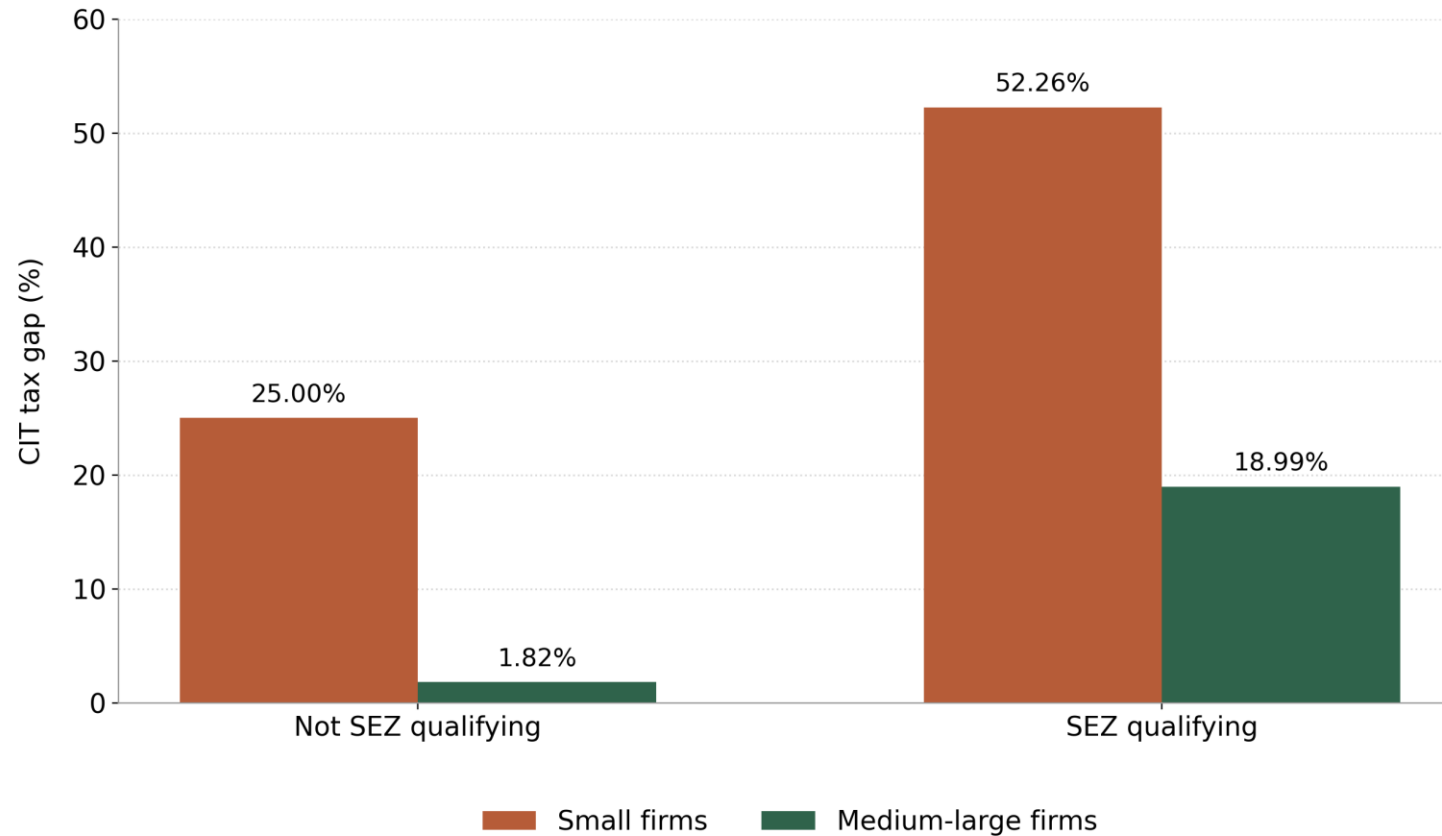
Tax gap: share of each group's potential CIT. Contribution: share of the medium-large firm gap.

SA-TIED

Southern Africa – Towards Inclusive Economic Development

CIT Compliance reporting gap

by SEZ status

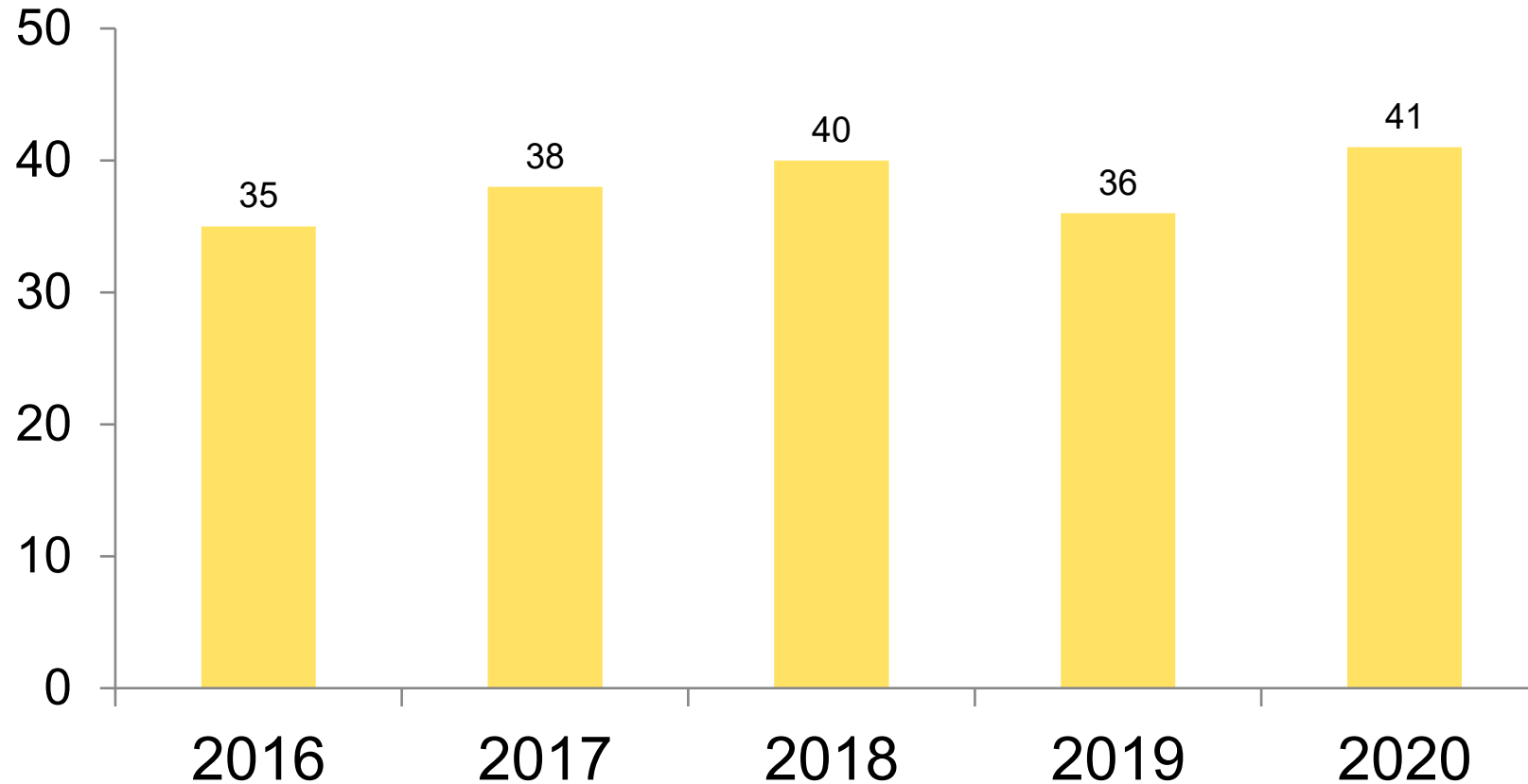


SA-TIED

Southern Africa – Towards Inclusive Economic Development

VAT Compliance reporting gap

By year

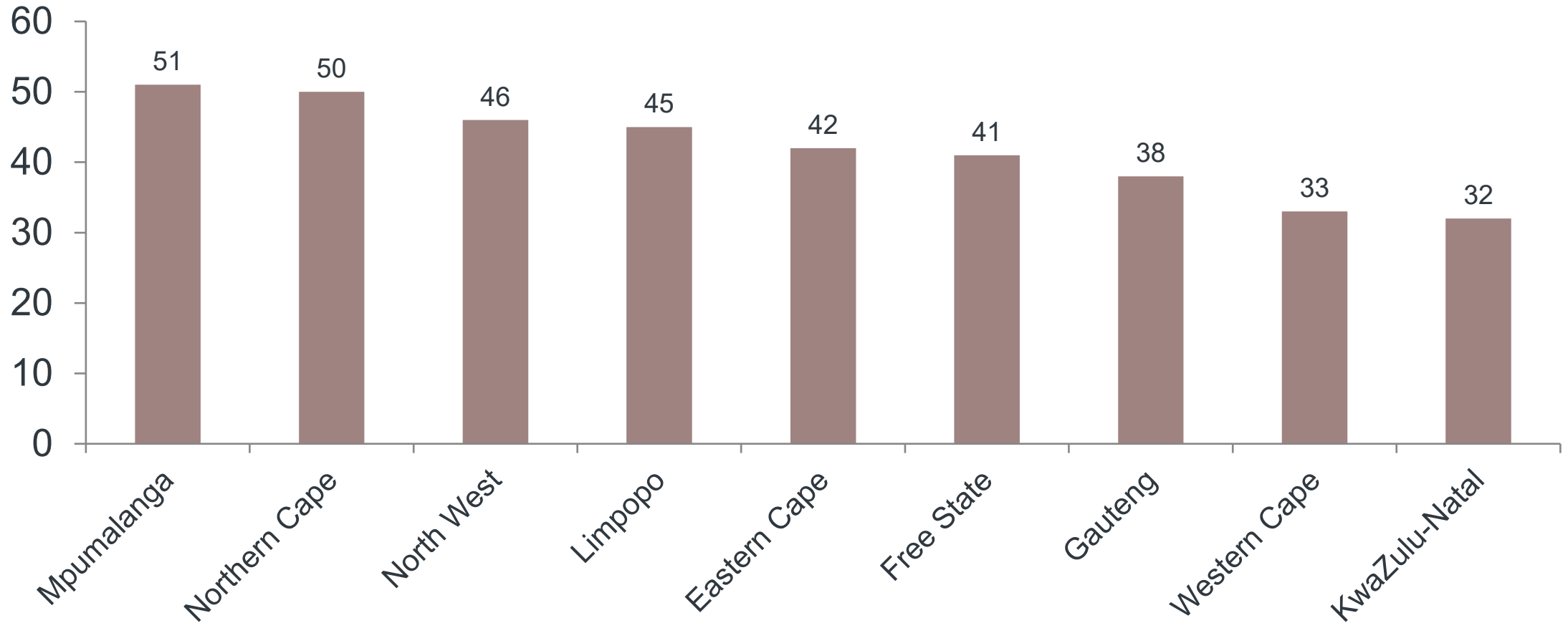


SA-TIED

Southern Africa – Towards Inclusive Economic Development

VAT Compliance reporting gap

by Region



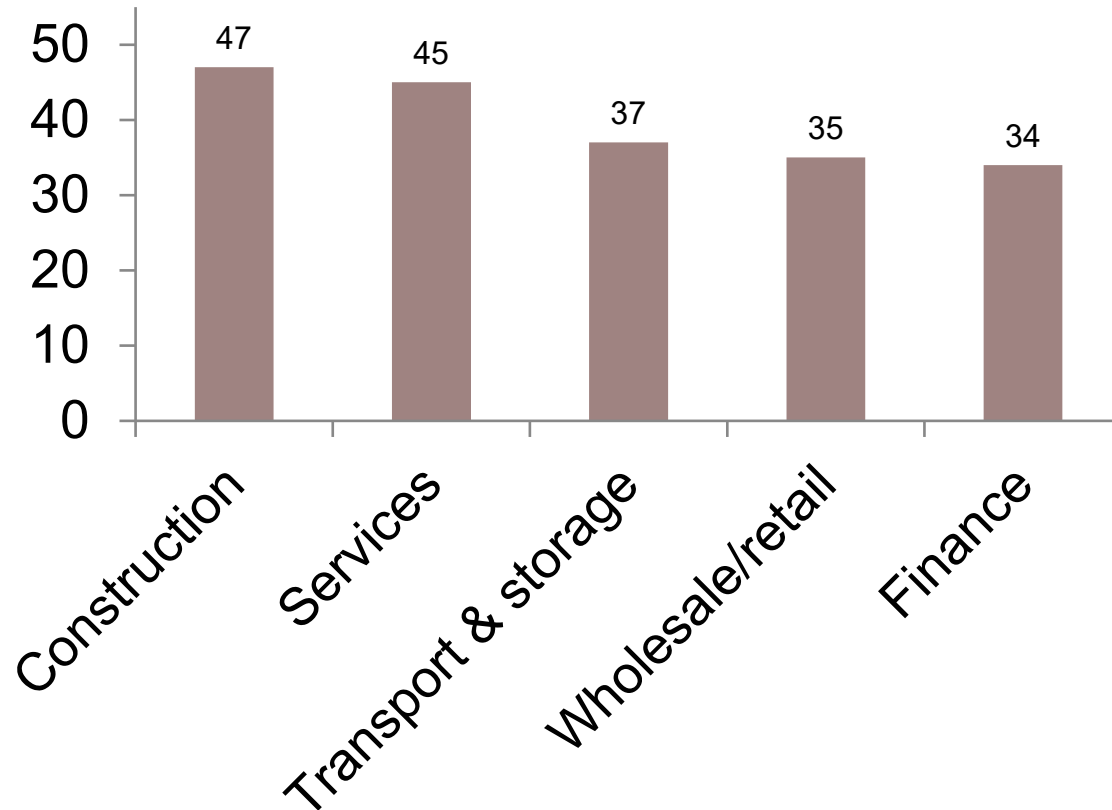
SA-TIED

Southern Africa – Towards Inclusive Economic Development

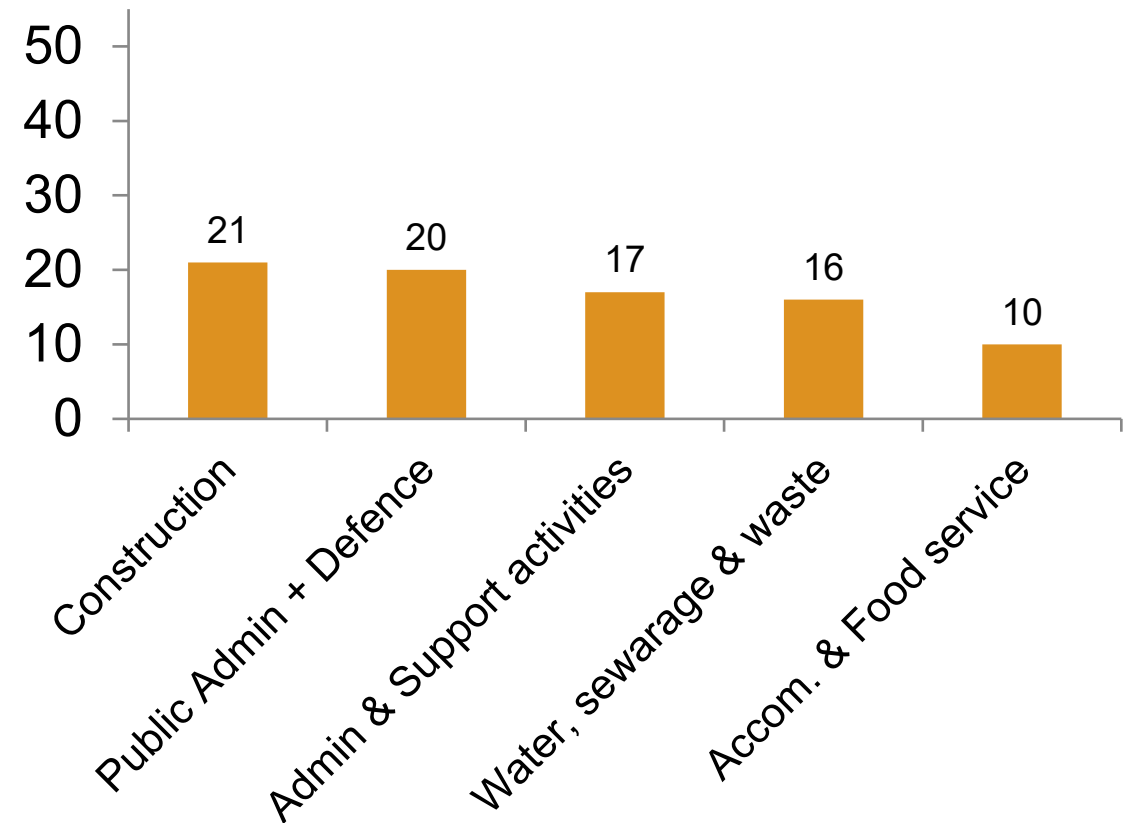
VAT & CIT Compliance reporting gap

by Industry (Top 5)

VAT Gap



CIT Gap



SA-TIED

Southern Africa – Towards Inclusive Economic Development

Summary

- A persistently large VAT compliance reporting gap over the 2016–2020 tax years, fluctuating between 35 and 41 per cent, with evidence of an upward trend.
- Sectoral and provincial disaggregation reveals that the VAT gap is *unevenly distributed*.
- Elevated VAT gaps are concentrated in **structurally high-risk sectors** such as construction, transport, and other services.
- The estimated CIT reporting gap averages **13.5%** of potential revenue over 2016–2020, with substantially higher gaps among small firms (around 25%) than medium-large firms (around 2%).
- The CIT gap is unevenly distributed across sectors and firm characteristics. **Construction** has the highest sectoral gap, while SEZ-qualifying firms show higher reporting gaps and multinationals lower gaps than their respective comparison groups.

Policy implications

Audit targeting

- The findings support differentiated audit strategies by firm type and sector.
- Certain sectors warrant sustained enforcement attention given their persistently high relative gaps.

Small firm compliance

- The high rates of misreporting among small firms call for measures that go beyond enforcement alone.
- Simplifying compliance requirements, providing clearer guidance on obligations, and making it easier for small businesses to report accurately can reduce inadvertent non-compliance alongside deliberate evasion.

SA-TIED

Southern Africa – Towards Inclusive Economic Development

Policy implications

Addressing MNE tax advantages through law, not enforcement

The relatively low underreporting rates among multinationals indicate that their tax advantage predominantly arises through legal or legally grey channels.

Addressing this requires policy and legislative reform.

Reconsider tax incentive programmes

The elevated evasion rates among SEZ-qualifying firms point to the extra costs of preferential tax regimes. Enforcement costs need to be taken into account in a cost-benefit assessment of incentive programmes.

Analysis next steps

- More segmentation of VAT vendors by **liability** (refund, zero liability, payable), **size** (small, medium, large)
- Trend analysis by taxpayer segments, with % contribution to the overall gap.
- Improving the **model**
 - Select the relevant variables (reduce overfitting)
 - Apply an updated two-step model to better model compliant taxpayers (zero assessment)
 - Extend to the XGBoost ML algorithm
 - Consider a split model for Assurance and Investigative audits
- Extend the time period of the analysis (**new data available**)

Conclusion

- Tax gap analysis helps us to understand the share of potential tax not collected; it is not a forecast of immediately recoverable cash.
- The challenges in data and methodologies mean point estimates have a degree of imprecision; the importance lies in the trend.
- Tax gap estimates should be redone regularly: additional audit years mature and recent target years become fully observable.
- Enforcement is expensive → focus on improving voluntary compliance
- Tax incentives as part of the policy gap should be evaluated against outcomes and costs.

Tax Gap research at UNU-WIDER

Start: Understand the Tax Gaps in Zambia, estimating VAT and CIT at disaggregated levels [[Working Paper](#), [Research Brief](#), [Journal article](#)]

Continue: Estimated the VAT Gap in Tanzania to improve tax compliance [[Working Paper](#), [Research Brief](#), [Journal Article](#)]

Launched ([Dec 2024](#)): A toolkit that is useful in other contexts for regular and consistent estimation

Aim: Equip Revenue Authorities with a machine learning framework to estimate tax gaps for VAT, PIT, and CIT, enhancing tax compliance and policy formulation.

Ongoing: Estimations in Uganda (CIT), Rwanda (VAT), and in 2027 in [Zimbabwe \(VAT\)](#), [Ethiopia](#) and [Benin](#).

Interest: Expressed from Eswatini, Sierra Leone, and own application in Indonesia

Next Phase: R/Python extensions, enhanced ML models

SA-TIED

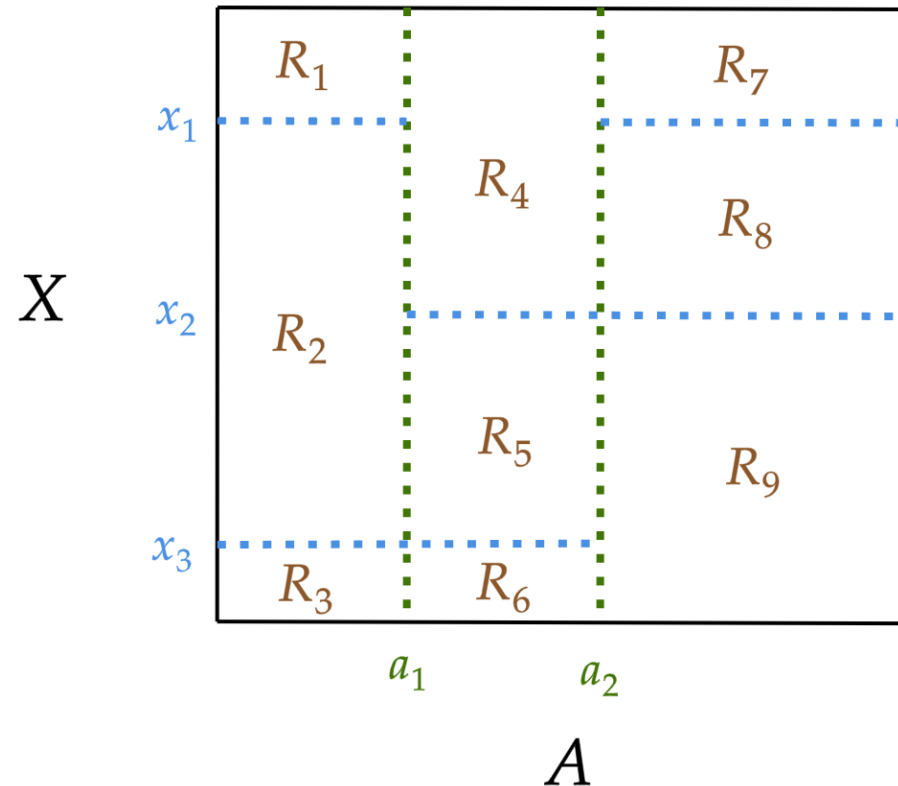
Southern Africa – Towards Inclusive Economic Development



SA-TIED

Southern Africa – Towards Inclusive Economic Development

Estimation of VAT gap - Random forest



- For a set of k covariates, the sample is split n times.
- In each split, the best prediction for the observation is estimated.
- Random forest uses bootstrapping to get resamples. To do this, only use part of the k covariates. This reduces the correlation across bootstrapped resamples.

Accounting for audit case selection

- Simple extrapolation to all taxpayers without accounting for selection would give a biased picture
- Audits are risk based:
 - => overestimation of the economy-wide gap
- But large firms are audited more often
 - They would typically be relatively more compliant
 - => underestimation of the economy-wide gap
- When risk-based audits are used for the bottom-up approach, key to deal with the selection