United Nations University
World Institute for Development Economics Research

Spotlight: Tax gap toolkit

UNU WIDER

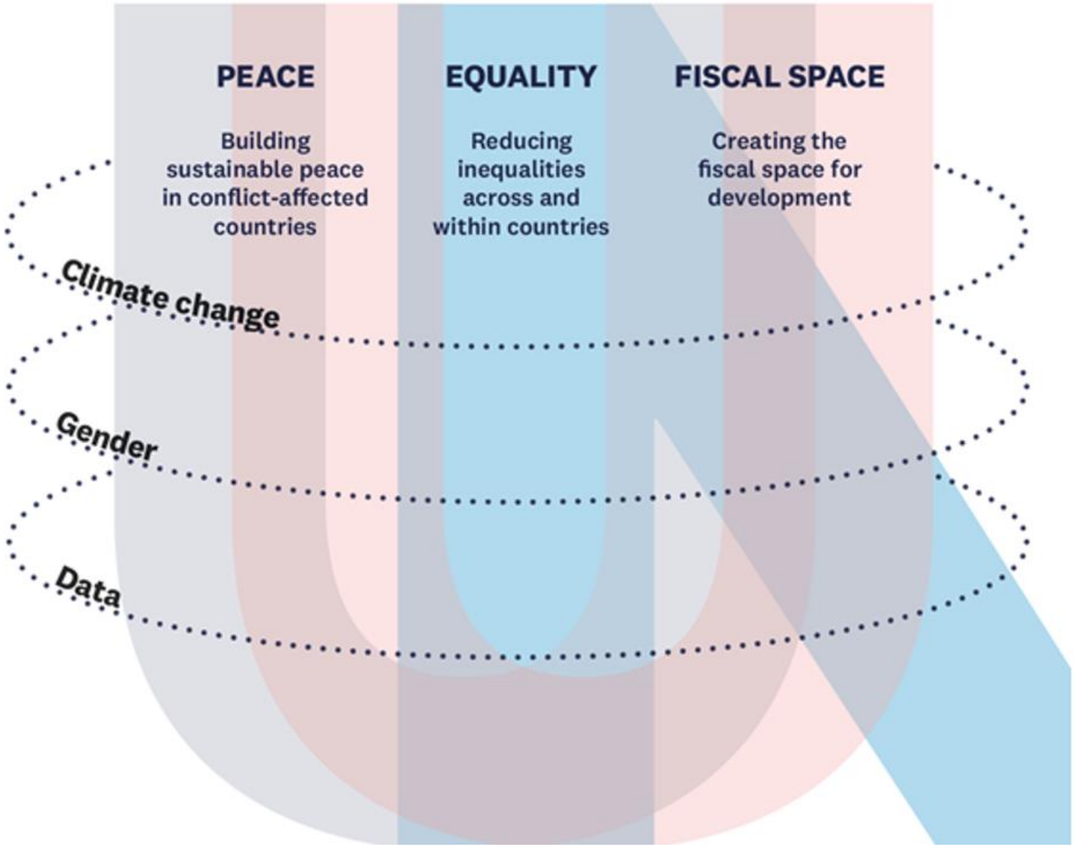
Amina Ebrahim, Research Fellow, UNU-WIDER

Spotlight session II | ATI General Assembly | 23 July 2025 | Addis Ababa

About: UNU-WIDER

- Core set of activities:
  1. Undertaking research & policy analysis
  2. Providing a forum for knowledge-sharing & debate
  3. Offering capacity development & training

- The main aim of UNU-WIDER's 2024-27 work programme is to harness research and evidence to address three mega challenges which will define the course of progress towards sustainable development in the decades to come:
  - **PEACE:** Building sustainable peace in conflict-affected countries
  - **EQUALITY:** Reducing persistent inequalities across and within countries
  - **FISCAL SPACE:** Creating the fiscal space for development.



Tax Research for Development

Main objectives:

1. **Strengthen tax administrators' skills** through research collaboration
2. **Help policymakers** create better development policies with research insights.
3. **Build communities** focused on using tax data in the Global South.

Tax Gap research

- Start: Understand the Tax Gaps in Zambia, estimating VAT and CIT at disaggregated levels
- Continue: Looking at VAT Gap in Tanzania to improve tax compliance
  - paper forthcoming in CESifo Economic Studies
- Develop: A toolkit that is useful in other contexts for regular estimation
- Ongoing: Estimations in Uganda, Rwanda and South Africa.

The Importance of the Tax Gap Estimation Toolkit

- The toolkit is a critical tool for identifying tax compliance gaps & designing evidence-based policies.
- It equips countries with tools to evaluate and explain shortfalls of tax revenues and address revenue losses effectively.
- It promotes transparency, accountability, and equity in tax systems and useful input for compliance risk management (CRM).
- Using **micro-level data**, the toolkit allows for granular estimation to identify ways to improve tax collection given revenue authority constraints.
- With its comprehensive and **adaptable framework**, the toolkit promotes regular and consistent estimation.

Developed as part of Commitment 1 of the ATI Declaration 2025 – *ATI partner countries commit to enhance DRM based on equitable tax policies as well as efficient, effective and transparent revenue administrations*.

Definition

The tax gap is the difference between the *actual* and *potential* tax collected under full compliance with the tax code.

Terminology:

- From a policy perspective: distinguish between the **compliance gap** and the **policy gap.**
  - **Compliance gap**: additional revenues raised if tax legislation perfectly enforced.
  - Policy gap: revenue that could be obtained (assuming perfect enforcement) if the base were taxed at the current standard (basic, average) tax rate.

Two broad approaches to tax gap estimation: top-down and bottom-up. Top-down techniques: Use independent statistical data, typically from national accounts, to model a potential tax base. Bottom-up techniques: Gross up or extrapolate from random or operational audit results.

Why estimate the tax gap?

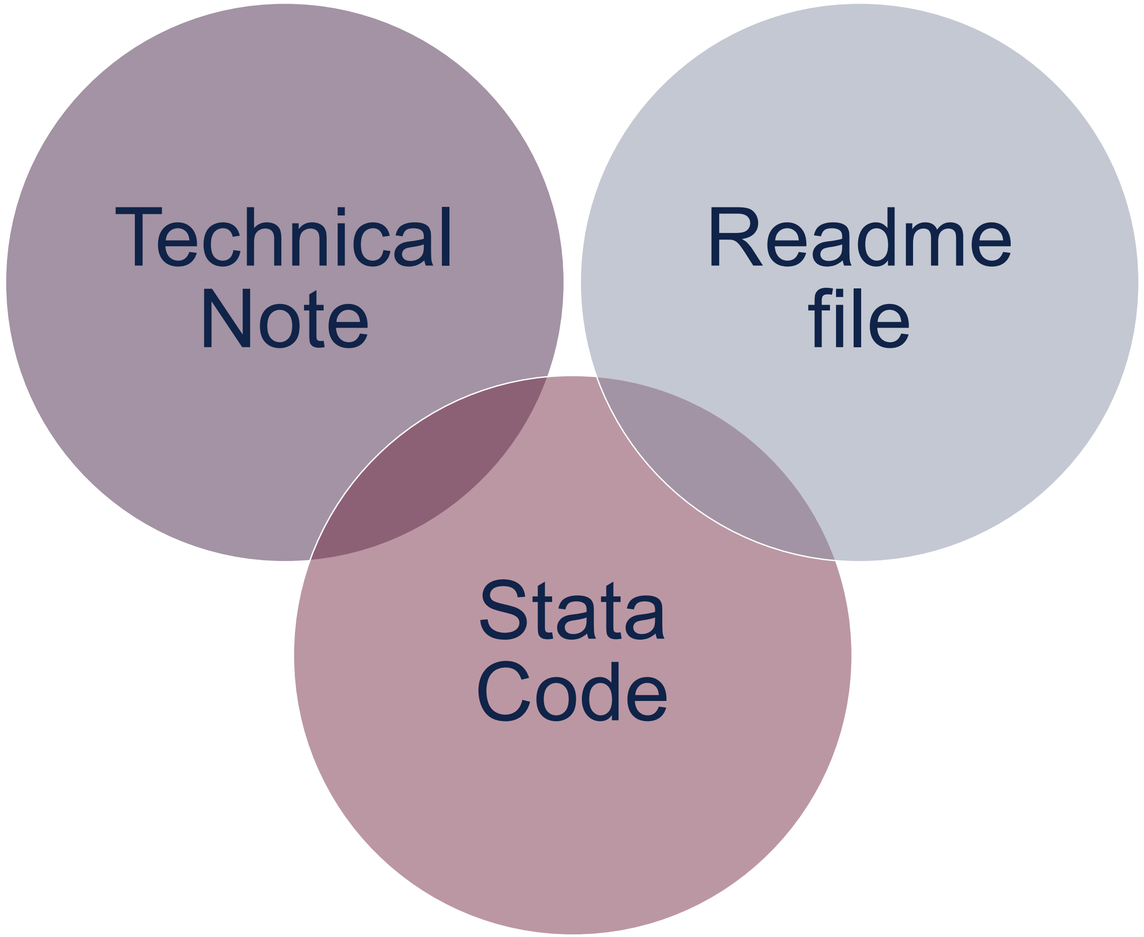
- As a high-level *performance indicator* of the tax administration's overall effectiveness in collecting tax revenues.
  - Complementing tax-to-GDP ratios as targets.

- To better explain shortfalls of tax revenues.
  - Tax gap analysis provides a sense of where the revenue leakage occurs.
  - It helps authorities implement corrective actions for improving revenue performance.

- Tax gap analysis is an important input for the tax compliance risk management (CRM) process.
  - Self-assessment systems require active monitoring of the extent of inaccurate reporting.

Source: "Estimating the tax gap – RA-GAP experience" by IMF Fiscal Affairs on ATI General Assembly June 22, 2023

What is the Tax Gap Toolkit?



A package of **adaptable materials** to implement tax compliance gap **estimation** based on the bottom-up approach using **machine learning**

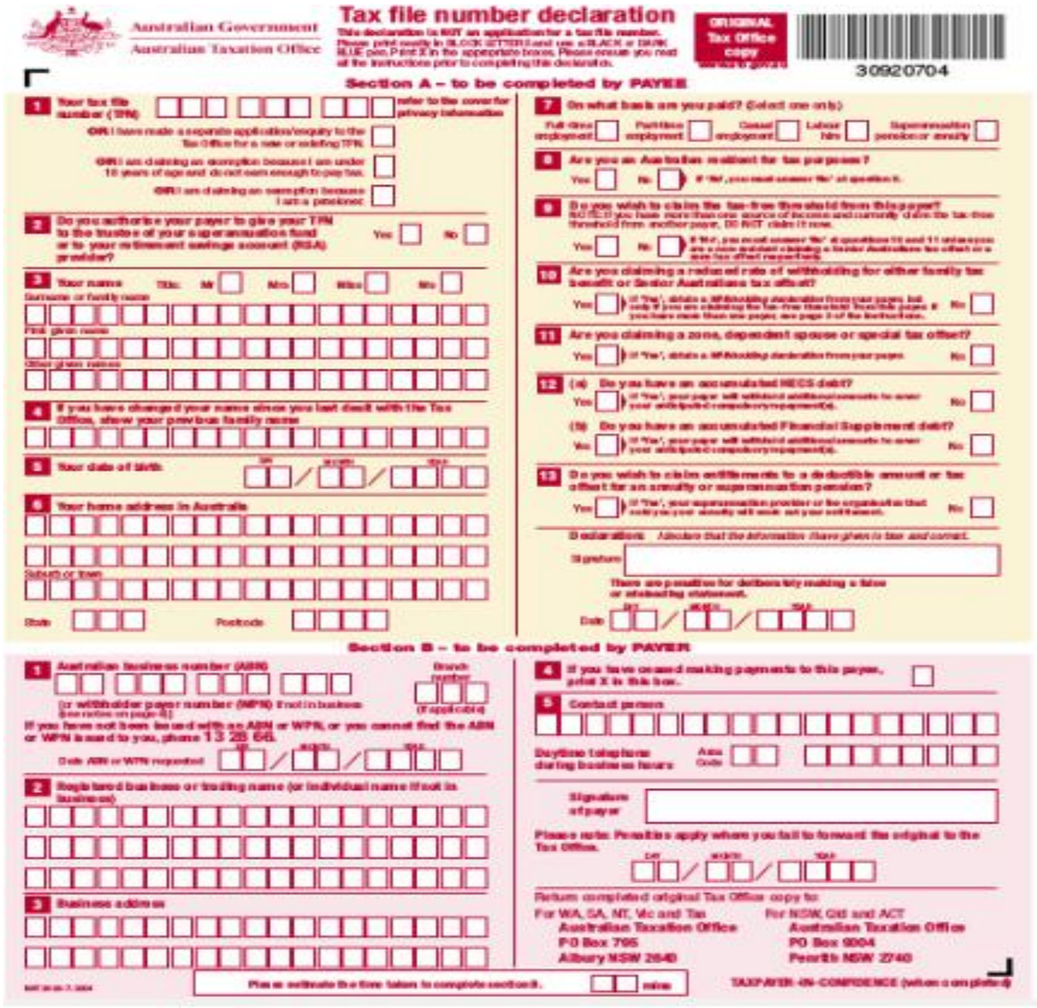
How does the Toolkit work?

Part 1 Data Preparation

Required data sources:

1. Tax returns: reported income, deductions, exemptions, & tax liability.
2. Audits: detected errors, adjustments, and penalties.
3. Supporting data: firm size, location, payments, industry.

- The toolkit guides the process of cleaning and preparing the data for analysis.


Australian Government
Australian Taxation Office

Tax file number declaration

ORIGINAL Tax Office copy

30920704

Section A – to be completed by PAYEE

Section B – to be completed by PAYER

How does the Toolkit work?

Part 2 Machine Learning Estimation

- The cleaned and prepared data are used to estimate the tax gap.
- For audited firms, we gather the unpaid/recovered tax information.
- The *challenge* with Tax Gap estimation that we don't know the potential tax for **unaudited firms** and unaudited periods.
- Toolkit uses info from audited firms to **estimate** the unpaid/potential recovered tax for unaudited firms and unaudited periods.
- Estimation is done using Machine Learning.

TPIN	AUDIT STATUS	RECOVERY ACTUAL	RECOVERY POTENTIAL
108469	Audited	17,158.95	
105476	Audited	14,921.44	
876560	Audited	14,118.49	
858323	Audited	16,138.63	
684517	Unaudited		-
124278	Unaudited		-
368696	Unaudited		-
335143	Unaudited		-
368696	Unaudited		-
212796	Unaudited		-
166410	Unaudited		-

How does the Toolkit work?

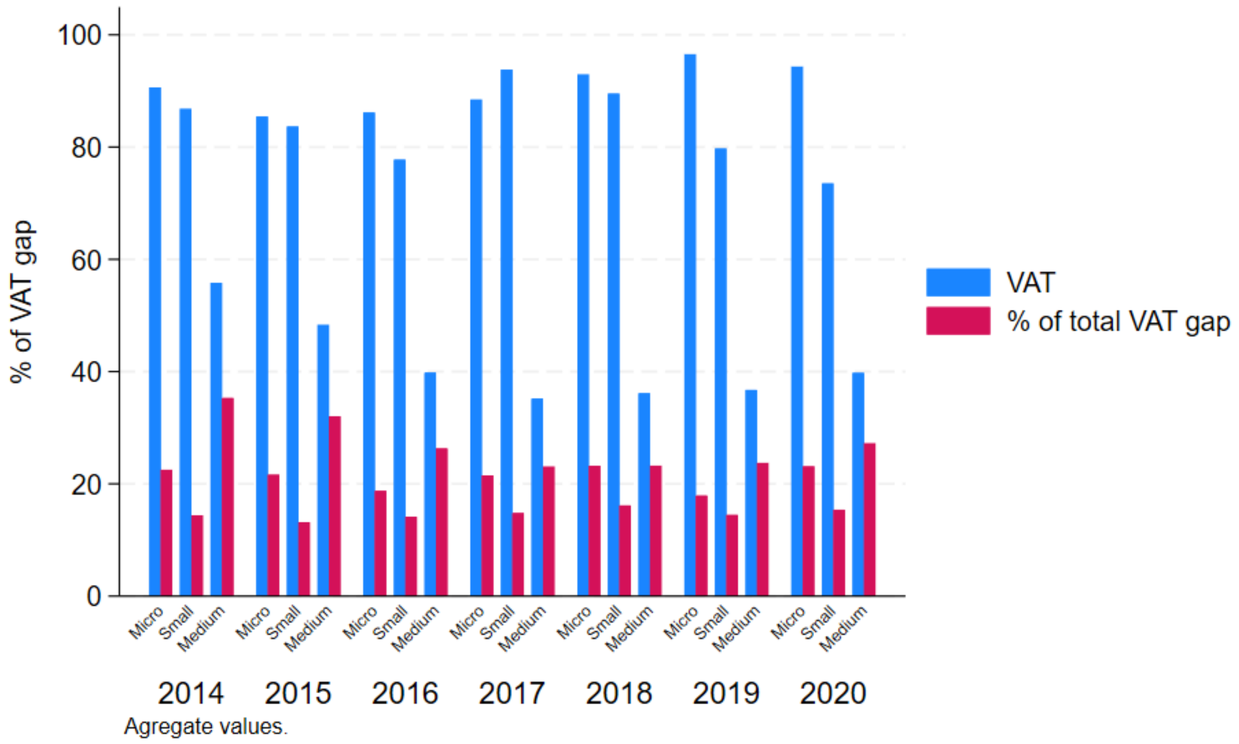
Part 2 Machine Learning Estimation

- The cleaned and prepared data are used to estimate the tax gap.
- For audited firms, we gather the unpaid/recovered tax information.
- The *challenge* with Tax Gap estimation that we don't know the potential tax for **unaudited firms** and unaudited periods.
- Toolkit uses info from audited firms to **estimate** the unpaid/potential recovered tax for unaudited firms and unaudited periods.
- Estimation is done using Machine Learning.

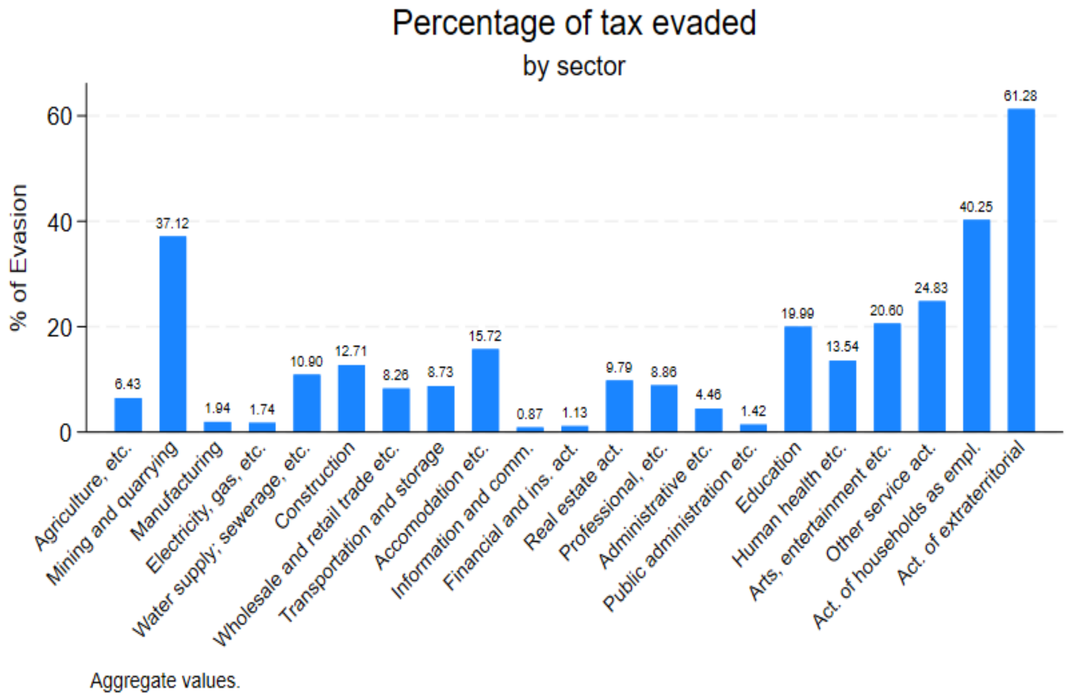
TPIN	AUDIT STATUS	RECOVERY ACTUAL	RECOVERY POTENTIAL
108469	Audited	17,158.95	
105476	Audited	14,921.44	
876560	Audited	14,118.49	
858323	Audited	16,138.63	
684517	Unaudited		6,517.63
124278	Unaudited		33,765.42
368696	Unaudited		22,256.54
335143	Unaudited		36,984.14
368696	Unaudited		16,175.93
212796	Unaudited		23,144.84
166410	Unaudited		13,103.30

Applications in ATI countries:

Tanzania

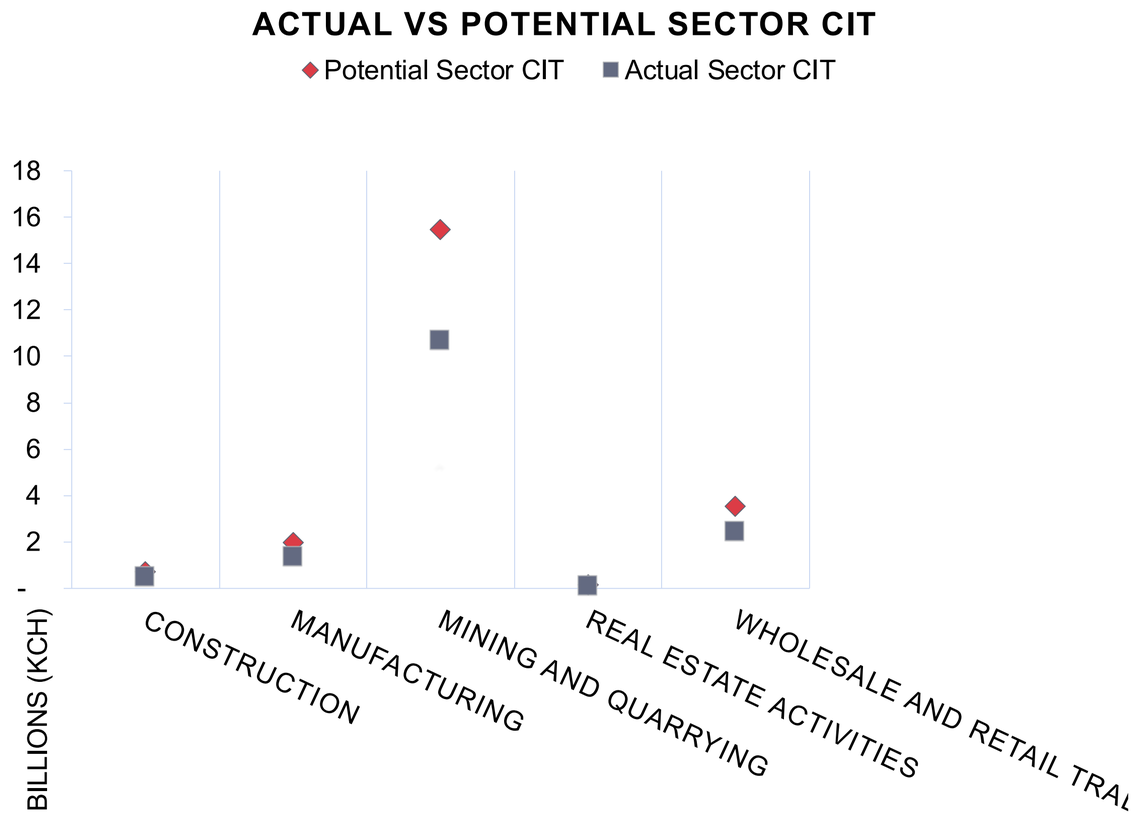


Uganda

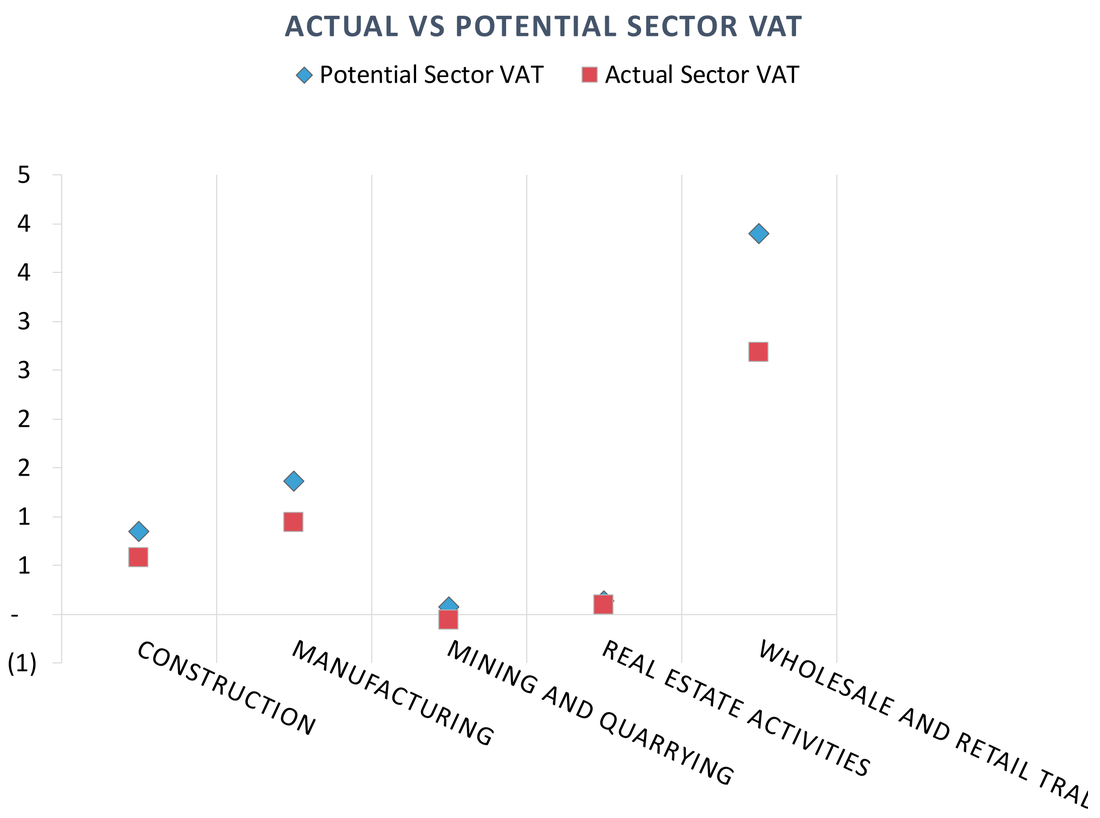


Applications in ATI countries: Zambia

CIT



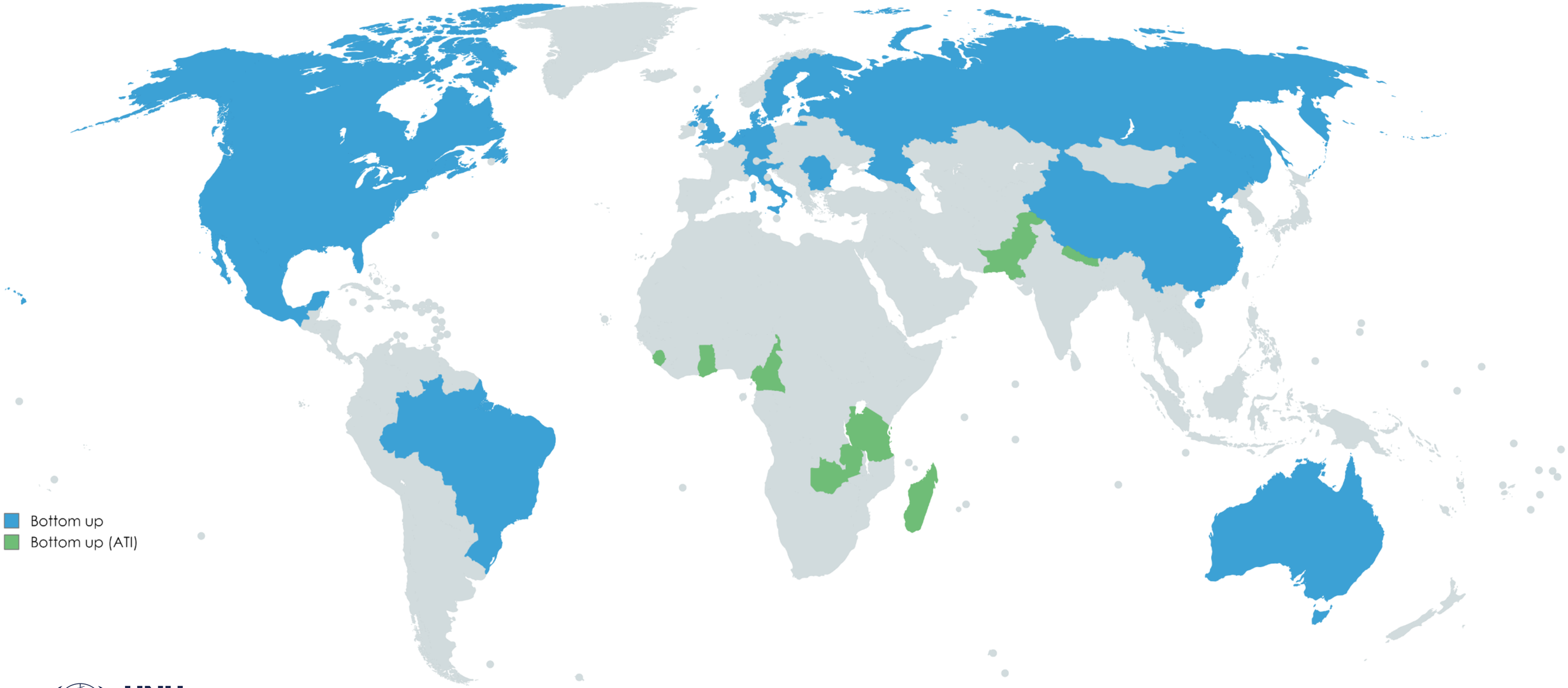
VAT



Key problems the Toolkit solves

- Gives countries options for tax gap estimation. Ideally, tax authorities should use more than one technique to cross-validate their tax gap results. The choice of modelling technique is also determined by the availability of data, time, and personnel.
- Launched in ready-to-use form, making it an asset for tax authorities wishing to embark on or improve their tax gap estimations.
- For ATI members, learning materials and training programmes are provided. This makes the model very accessible to tax administrations wanting to adopt it and institutionalise estimations.
- Great potential for further developments in other areas of tax compliance, such as the payments gap or the filing gap, and converting it to open source for broader adoption.

An achievable reality? Map of countries using a bottom-up approach



Created with mapchart.net

Feedback from the Uganda Revenue Authority:

Annual computation of the tax gap across tax heads and sectors can be done.
Target efforts to particular tax heads and sectors with higher tax gaps.
Results can be used as a proxy for effective tax risk management by tax head and sector.
Identify compliance issues as inputs to the compliance improvement plans.

A Call to Action

- Countries ready to implement can download the material for free!
- Stand ready to support the Toolkit’s implementation and explore collaborative opportunities.
- Support further application of the Toolkit in developing countries and invest in related capacity-building for sustainable, impactful results.
- Facilitate technical support and knowledge exchange to maximise its usability.

The Tax Gap toolkit was developed with support from the International Tax Compact (ITC), which facilitates the Secretariat of the ATI. The ITC is funded by the German Federal Ministry of Economic Cooperation and Development, co-funded by the European Union, and implemented by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.

UNU WIDER

Addis Tax Initiative

ati

german cooperation
DEUTSCHE ZUSAMMENARBEIT

Co-funded by the European Union

Facilitated by

itc

More information available online

wider.unu.edu

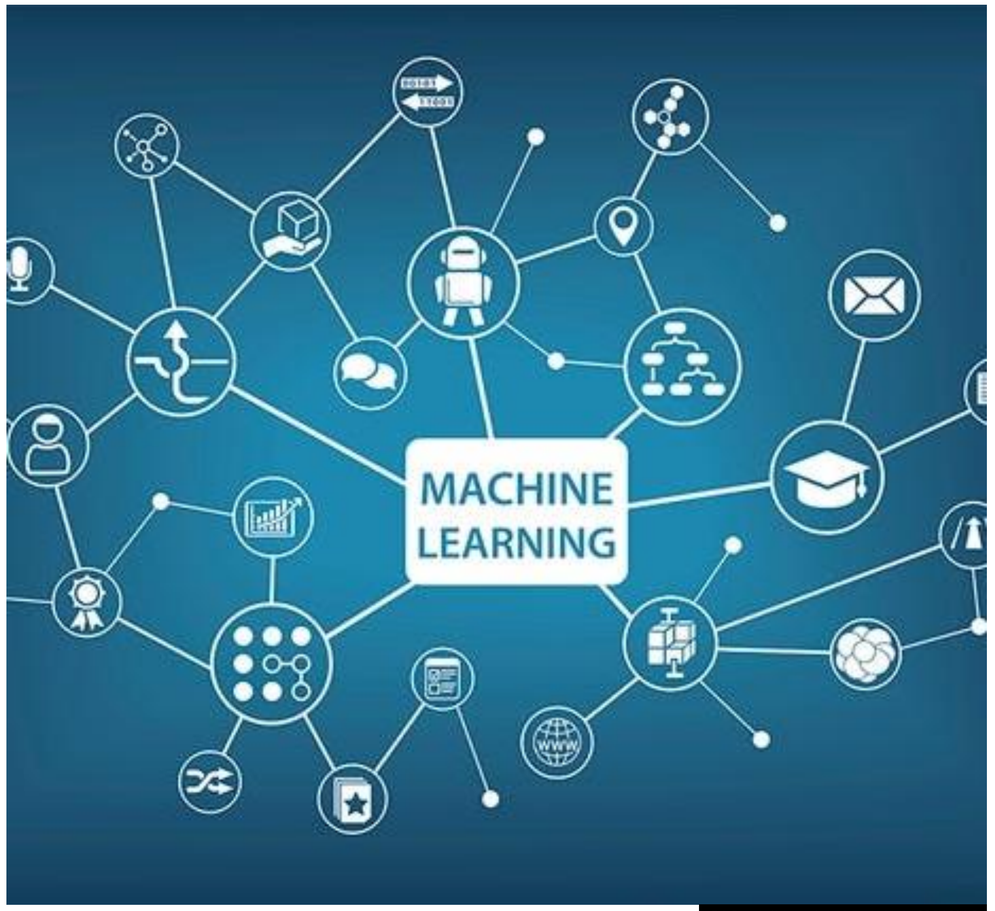
Follow us on LinkedIn, Facebook, and Twitter **@UNUWIDER**

Why machine learning?

Machine learning uses both *audited* and *unaudited* taxpayer data to identify patterns and learn the most relevant variables, then predicts the tax potential revenue.

Strengths of Machine Learning:

- Captures complex, non-linear relationships without strict distributional assumptions, and adapts to complex data structures.
- Improved prediction power over traditional statistical methods in predicting tax evasion.
- Addresses selection bias from observable factors by including all known variables influencing audit selection.
- Advanced ML techniques can incorporate methods, like the Heckman correction, to account for unobserved factors influencing selection bias.



This Photo by Unknown Author is licensed under CC BY-SA-NC