**United Nations University**
World Institute for Development Economics Research

# Bottom-up approach to Tax Gap estimation

**Amina Ebrahim**
**Research Fellow, UNU-WIDER**

2024 ATI Tax Gap Workshop | 19-21 March 2024 | Dar es Salaam, Tanzania

# Recap: The bottom-up tax (compliance) gap concept

- Revenue the government **should** collect minus the revenue the government **does** collect
- Knowing and understanding the **tax gap** helps the tax administration and finance ministries mobilise resources to reduce non-compliance.
- The main objective of this session is to describe the methods for the **bottom-up approach to tax gap** estimation.

# The Bottom-Up approach

Uses *micro-level* data from tax returns, audits, and risk registers to identify and quantify the components of the tax gap, such as under-reporting, under-payment, and non-filing.

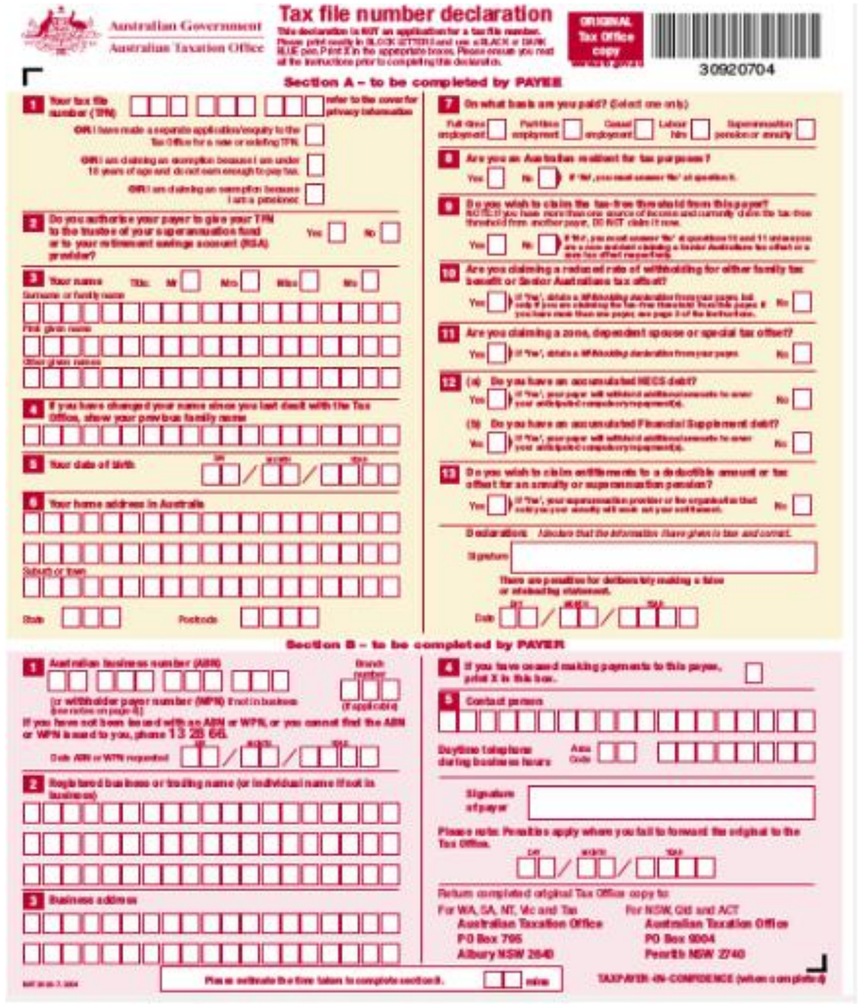
- **Advantages**: granularity, accuracy, and actionability,
- **Limitations**: data not digitized or available, coverage gaps, and resource requirements.
- **Key challenge**: how to estimate the *potential revenue*

  *Potential revenue* is not observed, and one needs to estimate it

# Data Collection: Finding the right data

- Main 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.
- Access to data: Survey results
  - Tax records: 15 countries digitized!
  - Audit: 99% risk-based, with assessment information.

**With all this data, bottom-up tax gap estimation is within reach!**



Australian Government
Australian Taxation Office

**Tax file number declaration**

ORIGINAL Tax Office copy

30920704

**Section A – to be completed by PAYEE**

1 Your tax file number (TFN)

2 Do you authorise your payer to give your TFN to the trustee of your superannuation fund or to your retirement savings account (RSA) provider?

3 Your name

4 If you have changed your name since you last dealt with the Tax Office, show your previous family name

5 Your date of birth

6 Your home address in Australia

7 On what basis are you paid? (Select one only)

8 Are you an Australian resident for tax purposes?

9 Do you wish to claim the tax-free threshold from this payer?

10 Are you claiming a reduced rate of withholding for either family tax benefit or Senior Australians tax offset?

11 Are you claiming a zone, dependent spouse or special tax offset?

12 (a) Do you have an accumulated HECS debt?

(b) Do you have an accumulated Financial Supplement debt?

13 Do you wish to claim entitlements to a deductible amount or tax offset for an annuity or superannuation pension?

Declaration: I declare that the information I have given is true and correct.

Signature

There are penalties for deliberately making a false or misleading statement.

Date

**Section B – to be completed by PAYER**

1 Australian business number (ABN)

2 Registered business or trading name (or individual name if not in business)

3 Business address

4 If you have ceased making payments to this payer, print X in this box.

5 Contact person

Daytime telephone during business hours

Signature of payer

Please note: Penalties apply where you fail to forward the original to the Tax Office.

Return completed original Tax Office copy to:

For WA, SA, NT, Vic and Tas
Australian Taxation Office
PO Box 795
Albury NSW 2640

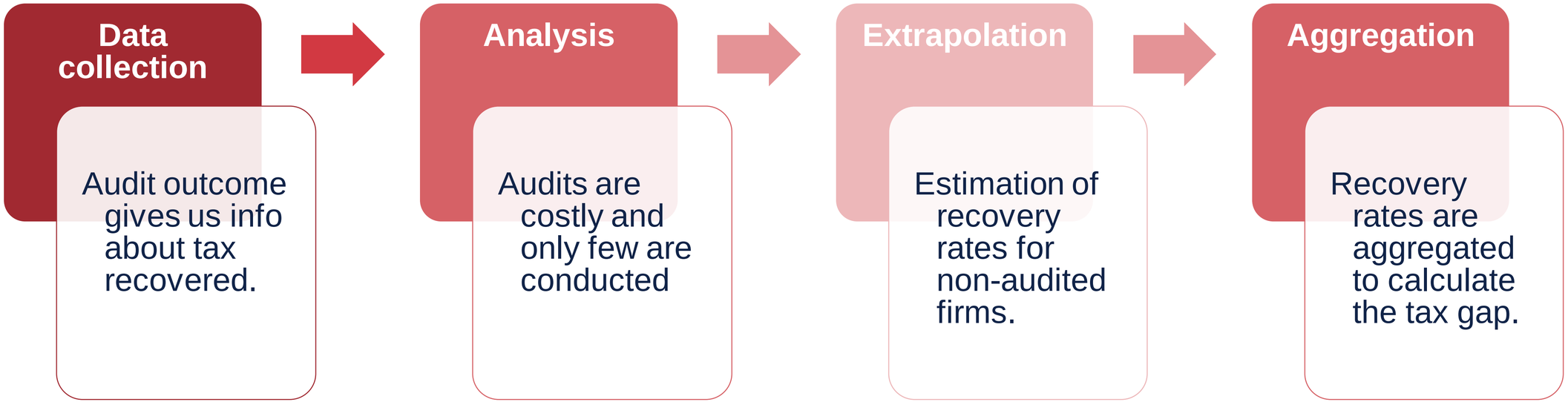
For NSW, Qld and ACT
Australian Taxation Office
PO Box 9004
Penrith NSW 2740

Please estimate the time taken to complete section B. mins

TAXPAYER-IN-CONFIDENCE (when completed)

# Data Collection: Benefits of Audit data

- Audit data can be more *sensitive* and harder to reach within institutions
  - 5 Safes: Safe Projects, Safe People, Safe Settings, Safe Data and Safe Outputs
- The key part of the audit information needed is the **tax recovered** and the audit **period**.



# Analysis: Bottom-up methodology

- **Context** of tax system
  - What are the rules and regulations of tax system?
  - Are there exemptions that need to be considered?

- Analysis depends on the **audit selection** used.
  - Random audits: provide an unbiased population estimate if truly random.
  - Risk-based audits: techniques must be used to infer the recovery rate for all taxpayers

# Extrapolation: methodology options

- Various methods of *extrapolation* are used for the bottom-up approach:
  - Regression: estimates the relationship between the tax gap and the chosen explanatory variables, such as income, deductions, and risk scores.
  - Machine learning: uses all available information to learn the most relevant variables, then predicts the tax potential revenue.
  - Extreme Value (EV): Based on fitting the distribution of the prediction to the actual one. Usually, a Pareto distribution is used where the Pareo parameter needs to be estimated.
  - Heckman (two-step) method: Estimates the distribution of the audit probability in the first stage, which is later considered to estimate the structural model. Through this, the bias in selecting firms for auditing is considered in the prediction model.
  - Stochastic frontier model: Estimates a Cobb–Douglas production function. The gap is estimated as the distance between the actual revenue level and the “frontier” (potential revenue)
- Each method provides different levels of accuracy and precision for the tax gap depending on the context (available data and tax rules)

# Extrapolation: Machine learning

| TPIN | Audit Status | Recovery Actual |
|---|---|---|
| 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 | |

# Extrapolation: 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 | | |

# Extrapolation: 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 |

# Aggregation: Tax compliance gap for CIT, PIT, VAT and more

- Summation: involves adding up the tax gap estimates from the different data points and components to obtain the total tax gap for the tax type.

  - Sum over the tax year
  - Sum over the industry
  - Sum over firm size

# Usefulness:

- This work for other tax types: trade and excise taxes
- Application to CIT, VAT and PIT
- Decomposition for policy analysis
  - Small, medium, large firms
  - Firm behaviour
  - Sectoral analysis
  - Gross vs net gap

## Strengths

- More granular and actionable information on the nature and extent of non-compliance and the impact of the tax policy.
- Allows for cross-checking the results from the top-down approach and identifying potential errors in the data sources or methodologies.
- Enhances the transparency and credibility of the tax gap estimates

## Weakness

- Requires a large amount of data and resources and may not cover all the sources of non-compliance.
- Can take time to estimate and needs skilled persons in administration to collect and prepare the data.
- Audit selection needs to be carefully modelled, needs to know the context and practices well.

# Summary

- The bottom-up approach to tax gap estimation uses micro-level data from tax returns, audits, risk registers to identify and quantify the tax gap.
- Caution in interpretation – point estimates make less sense than trends over time.
- What is the ideal tax gap? Good to know what neighbouring countries are getting but tax systems are different, and methods may vary.
- Advantages include: granularity, accuracy, and actionability, disadvantages include: data limitations, coverage gaps, and resource requirements.
- Informality not covered – need surveys to gauge the tax base and the likely revenue in firms outside of the tax net.
- The bottom-up approach can complement the top-down approach by providing cross-checks and insights into the sources and causes of non-compliance, and informing the decision-making and resource allocation of the tax administration.

More information available online

# wider.unu.edu

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