

TAX GAP ESTIMATION TOOLKIT TRAINING

Programme

Day 1: Monday, 7 October 2024

- 14:00 – 15:00 Opening remarks
Introduction to the Bottom-Up approach with Q&A
- 15:00 – 15:15 Short Break (15 minutes)
- 15:15 – 16:00 Data Preparation theory
- 16:00 – 17:00 Data Preparation practical with Q&A – Group break-out session

Day 2: Tuesday, 8 October 2024

- 14:00 – 15:00 Estimation theory
- 15:00 – 15:15 Short break (15 minutes)
- 15:15 – 16:00 Estimation practical with Q&A – Group break-out session
- 16:00 – 17:00 Interpreting and implementing findings with Q&A
Closing words

Introduction to Tax Gap estimation - Bottom-up approach



UNU
WIDER

Dr Amina Ebrahim
Research Fellow, UNU-WIDER

2024 ATI Tax Gap Estimation Toolkit Training | 7 October 2024

Introduction to the team

Dr Amina
Ebrahim

Dr Maria Jousté

Dr Kwabena
Adu-Ababio

Prof. Jukka
Pirttilä

Dr Sebastián
Castillo

Mostafa Bahbah

Gerald Agaba

Hilja-Maria
Takala

Towards the fulfillment of ATI commitment 1

The analysis of the tax gap is relevant for the fulfilment of ATI Commitment 1 outlined in the **ATI Declaration 2025**, which points out that ATI members

“will enhance the effectiveness of partner countries’ revenue administrations in curbing non-compliant behaviour by strengthening their capacities and capabilities, including risk management frameworks.”



Why estimate the tax gap?

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

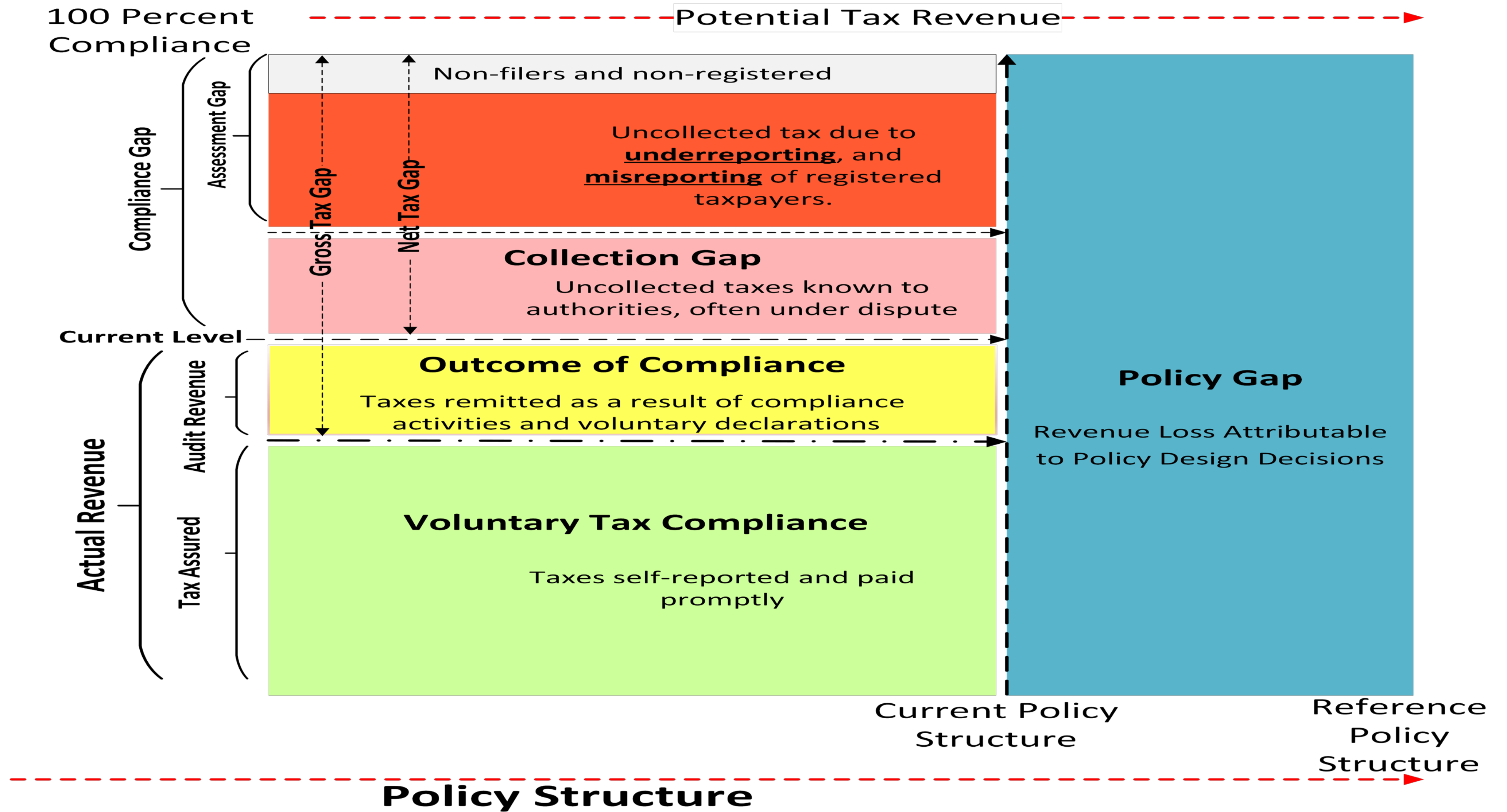
A few definitions

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.
- From a collection perspective: distinguish between gross gap and net gap
 - Non-filing, underreporting and underpayment

Level of compliance



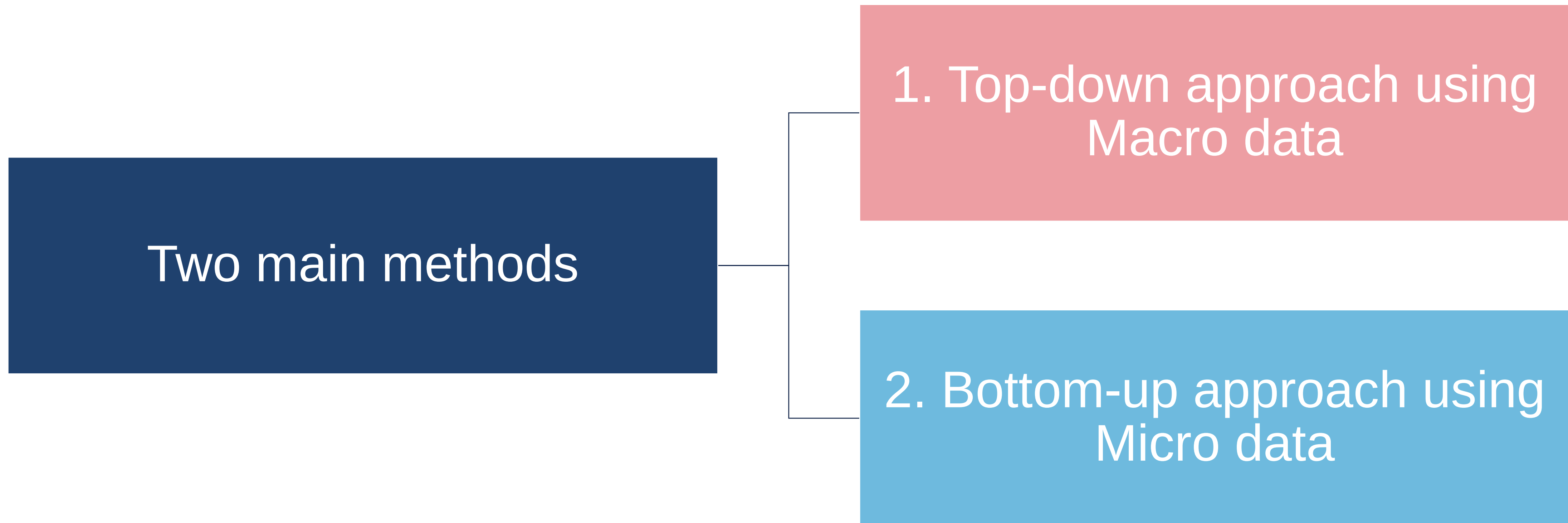
Tax compliance gap components

- **Non-payment component:** the difference between potential and actual tax revenue, reflecting the fraction of tax evaded through non-reporting or under-reporting to the tax authority.
- **Non-filing component:** the difference between the number of potential and actual tax filers. It shows the proportion of entities that should have filed returns but did not, highlighting the part of the tax base that escapes taxation due to non-reporting.
- **Under-reporting component:** the difference between a taxpayer's true tax liability and the reported amount. This occurs when taxpayers report less income than they actually earned or claim more deductions, credits, or other tax benefits than allowed by law, or a combination of both.

Important considerations

- Estimates of tax gaps are *rough* indicators of revenue loss.
- Usefulness of tax gap estimates depends on the methodology and the **data** used to estimate the gaps.
- **Caution** in interpretation needed –tax gap estimates are imprecise. It is, therefore, best to emphasise trends in the estimated results instead of focusing on absolute numbers.
- Differences in tax systems and tax gap methods mean that comparisons between countries are limited.
- Many of the factors driving changes in the tax gap are outside of the control of the tax administration

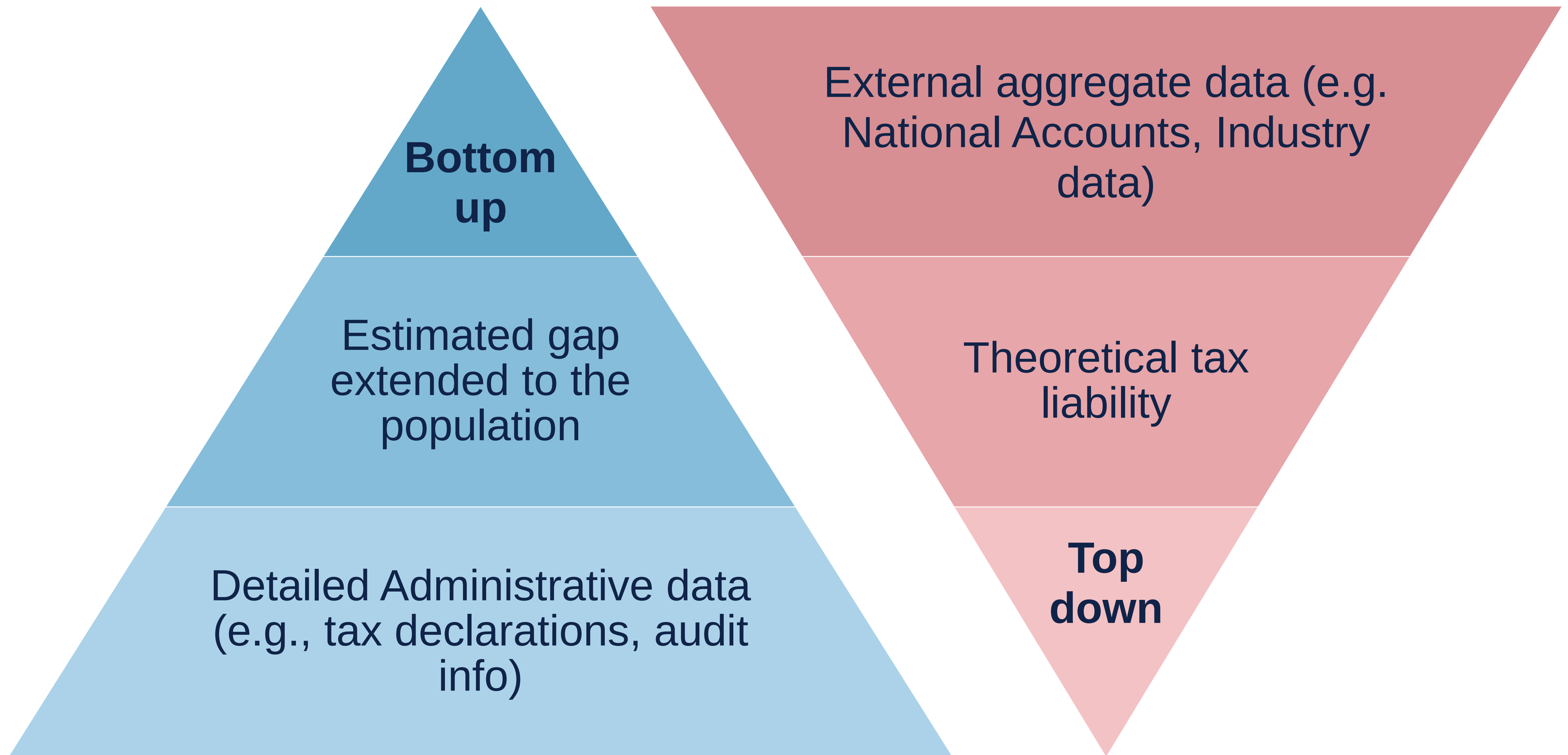
Methodology



Which method to choose?

Considerations for suitability of method:

1. The structure of the tax system
 2. Potential areas of compliance risks
 3. Available data and resources
- For quality assurance, attempt multiple approaches (i.e. a **top-down** approach and a **bottom-up approach**)



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: Earlier 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!

The image shows a sample of an Australian Tax file number declaration form. It is divided into two main sections: Section A (to be completed by PAYEE) and Section B (to be completed by PAYER).

Section A - to be completed by PAYEE

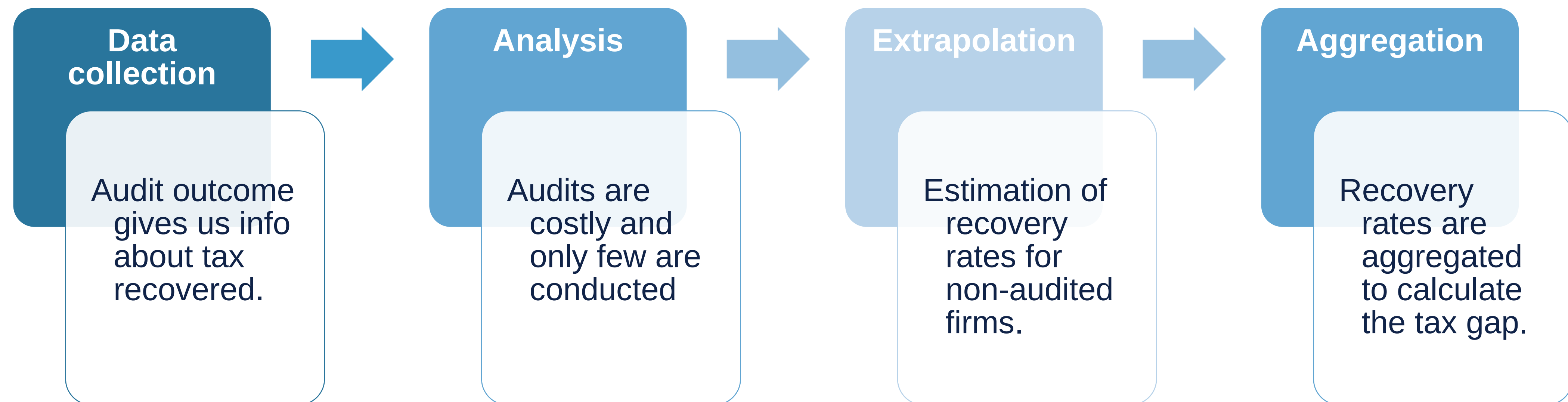
- 1. Your tax file number (TFN):** A box for the TFN, with a note that it must be a separate application to the ATO.
- 2. Do you authorize your payer to give your TFN to the trustee of your superannuation fund or to your retirement savings account (RIS)?** Yes/No options.
- 3. Your name:** Surname or family name, First given name, and Other given names.
- 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:** Day, month, and year.
- 6. Your home address in Australia:** Street or town, State, and Postcode.
- 7. On what basis are you paid?** (Select one only): Full-time employment, Part-time employment, Casual employment, Labour hire, Superannuation pensioner, or Annuity.
- 8. Are you an Australian resident for tax purposes?** Yes/No.
- 9. Do you wish to claim the tax-free threshold from this payer?** Yes/No.
- 10. Are you claiming a reduced rate of withholding for either family tax benefit or Section 80-108 tax offset?** Yes/No.
- 11. Are you claiming a 20% dependent spouse or special tax offset?** Yes/No.
- 12. (a) Do you have an accumulated HECS debt?** Yes/No. (b) Do you have an accumulated Financial Supplement debt? Yes/No.
- 13. Do you wish to claim offsets for a deductible amount or tax offset for an annuity or superannuation pension?** Yes/No.

Section B - to be completed by PAYER

- 1. Australian business number (ABN):** A box for the ABN, with a note that it must be a separate application to the ATO.
- 2. Single trade business or trading name (or individual name if not in business):** A box for the name.
- 3. Business address:** A box for the address.
- 4. If you have ceased making payments to this payer, print 2 in this box.**
- 5. Contact person:** Daytime, evening, and mobile phone numbers, and email address.
- Signature of payer:** A box for the signature.
- 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, P.O. Box 795, Albany NSW 2540. For NSW, Qld and ACT: Australian Taxation Office, P.O. Box 9994, Perth NSW 2040.

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. These regression techniques include the Heckman approach and the propensity score matching approach.
 - 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.
 - Key Considerations:
 - Powerful selection equation needed.
 - Requires at least one variable influencing audit selection but not non-compliance.
 - Propensity Score Matching (PSM): corrects selection bias by assigning propensity scores, which estimate the likelihood of audit, then matching audited and non-audited taxpayers with similar scores to impute values for non-audited entities based on the audited data.
 - Key Considerations:
 - Matching methods include nearest neighbor, kernel, etc.
 - Limited in accounting for unobservable factors.

Extrapolation: methodology options

- Extreme Value (EV): estimates the tax gap by focusing on extreme non-compliant taxpayers, assuming that a few significant cases (e.g., major tax evaders) disproportionately contribute to the gap, typically using a Pareto distribution.
 - Key Considerations:
 - Relies on the validity of the Pareto distribution, which may not always fit the data.
 - May overlook broader non-compliance trends by concentrating only on the most extreme cases.

Clustering : groups taxpayers into clusters based on key variables such as company size, geographical region, and industry sector. Within each cluster, scaling factors are applied to the audit results of audited taxpayers to extrapolate and estimate the tax gap for the entire cluster.

- Key Considerations:
 - The method assumes taxpayers within each cluster are similar based on selected variables (e.g., size, industry), but significant variations within clusters can lead to imprecise estimates if important differences are overlooked.

Extrapolation: methodology options

- 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.
- Outperforms traditional statistical methods in predicting tax evasion and can provide more reliable and representative tax gap estimates.
- 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.

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

Benefits and Limitations of the Bottom-Up approach

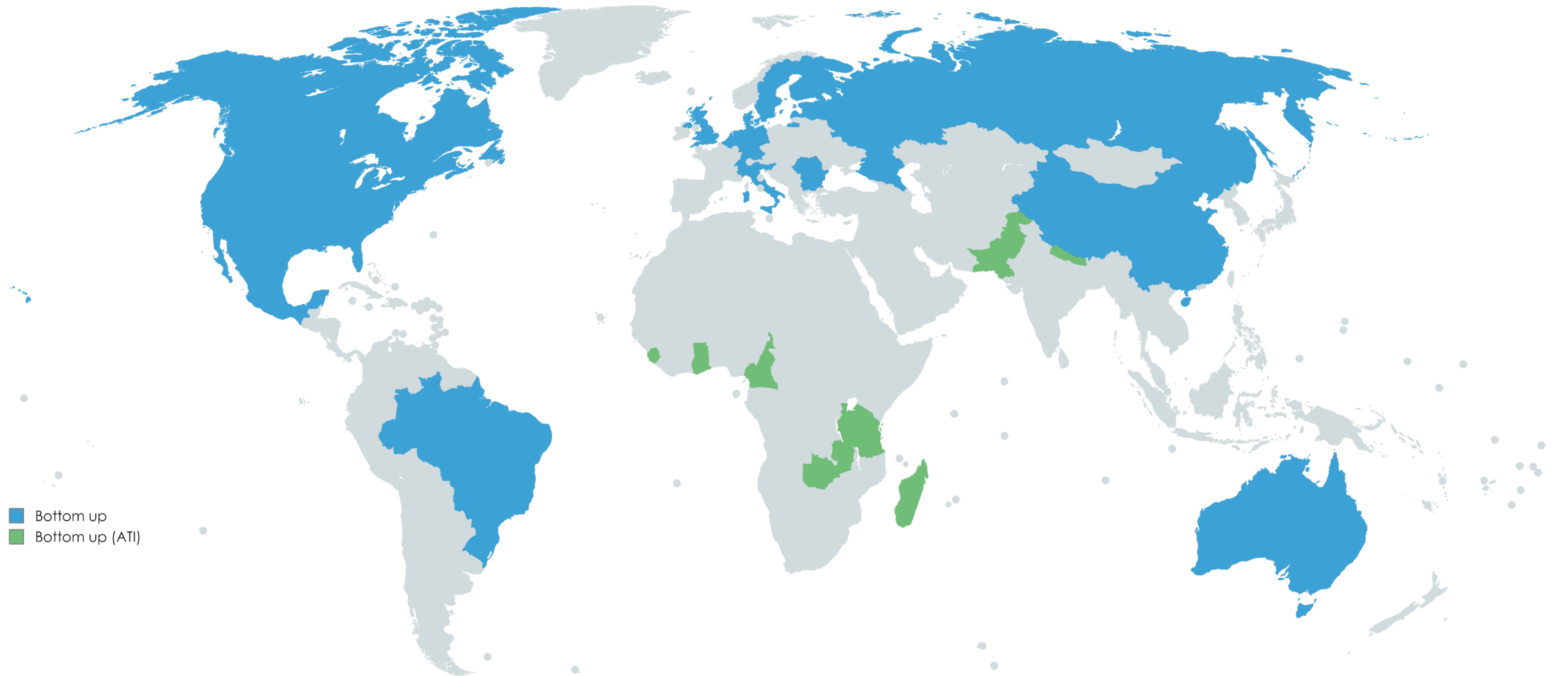


- Tax gap components are estimated separately for **different taxpayer groups** as well as types of non-compliance
 - For example: small, medium or large firms, specific sectors etc.
- Uses tax administrative data that is usually more **comprehensive** and based on individual cases
- Few **assumptions** to estimate the tax base.



- **Concealed** activities may be excluded
- Data intensive
- Audit selection process **complex to model**, and estimates sensitive to modelling assumptions

Reality? Map of countries using bottom-up approach



Summary

- The bottom-up approach to tax gap estimation uses micro-level data from tax returns, audits, and risk registers to identify and quantify the tax gap.
- **Caution in interpretation:** Point estimates are less meaningful than trends observed over time.
- **What is the ideal tax gap?** It is useful to know what neighbouring countries are achieving, but comparisons should be made cautiously, considering differences in tax systems and tax gap methodologies.
- **Advantages include** granularity, accuracy, and actionability. **Disadvantages include** data limitations, coverage gaps, and resource requirements.
- **Informality not covered:** Surveys are needed to gauge the tax base and likely revenue in firms outside of the tax net.
- The bottom-up approach can complement the top-down approach by providing granular cross-checks, insights into non-compliance, and informing tax administration decisions. This includes advanced extrapolation techniques like regression and machine learning, which can provide more accurate predictions of the tax gap.

Bottom-up approach materials

- This workshop trains to use the Bottom-up approach. Providing technical sessions with examples of estimating the **Tax gap** using **dummy data**
- Tax Gap Toolkit **Technical Note**, **Readme files** and example **Stata codes** will be available on the UNU-WIDER's webpage later this year



Thank you!

Amina Ebrahim

amina@wider.unu.edu



**UNU
WIDER**

More information available online

wider.unu.edu

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