

Practical session: Tax Gap estimation



UNU
WIDER

Break out groups:

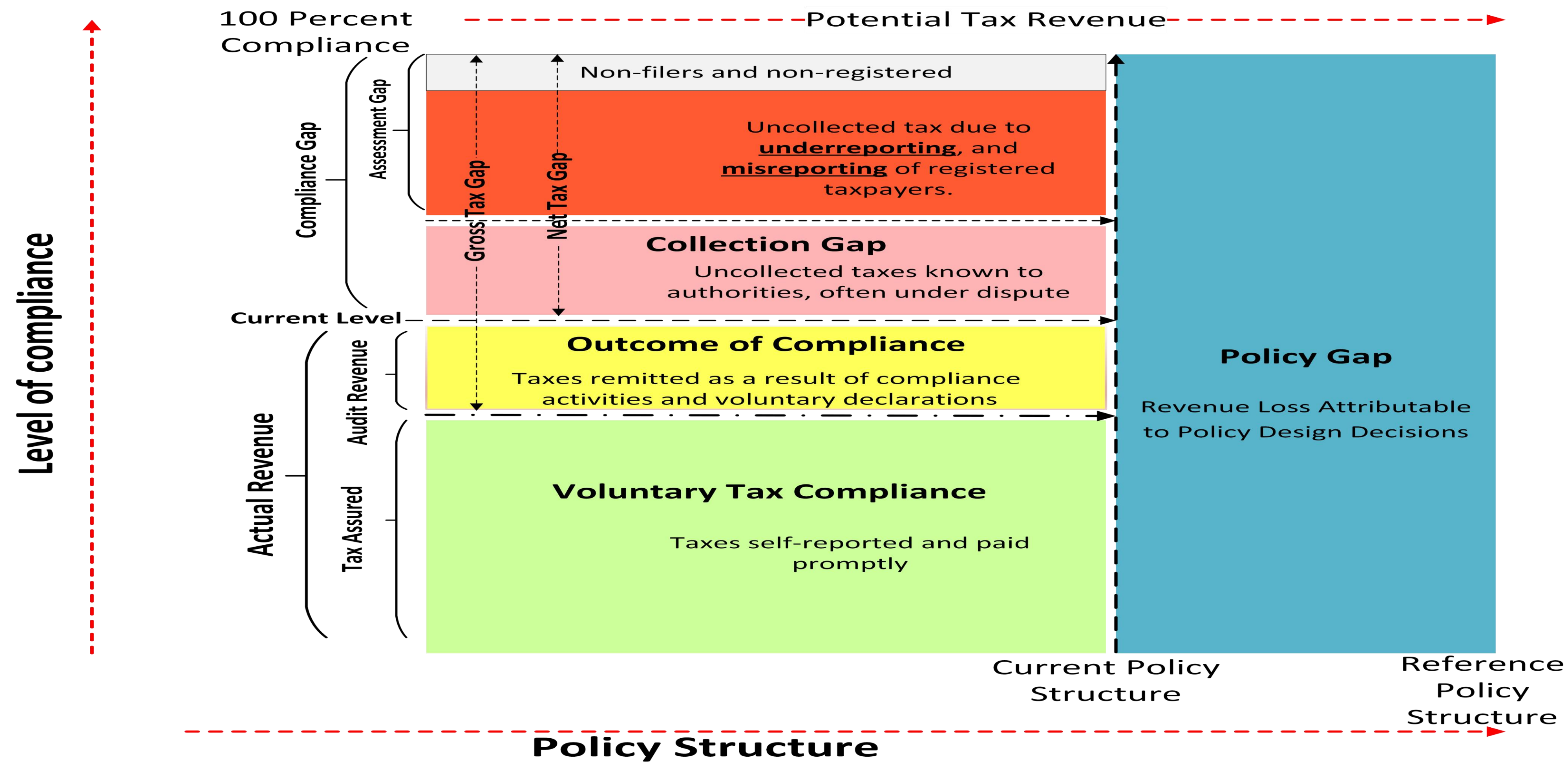
Maria Jouste + Gerald Agaba

Amina Ebrahim + Mostafa Bahbah

Kwabena Adu-Ababio + Hilja-Maria

2024 ATI Tax Gap Estimation Toolkit Training | October 7-8 2024

Recap of estimation goal



Recap of steps yesterday

- What did we achieve yesterday?
- 1. We covered the concepts of BU approach
- 2. Theory of Data cleaning
- 3. Collecting preparing data
 - Importing data to Stata
 - Dealing with Duplicates
 - Merging Audit and return data.

Aim of this session

1. Show/Demonstrate/Teach the steps towards the estimation – Toolkit Part 2.
2. Show how to investigate disaggregated results (by sector).
3. Show how to investigate disaggregated results (by return year).
4. Show how to investigate the effects/consequence of aggregate tax gaps.

We are not teaching how to use Stata merely guiding the steps to follow.

Own implementation requires contextualizing.

Preamble and Preparatory Steps

In the following exercises we will use the Machine Learning code to estimate tax gaps for VAT CIT and PAYE. For practical reasons we focus on only one tax type. However we encourage you to practice with the other tax types and use other variables to fully understand the usefulness of the toolkit. Before engaging in the preparatory steps run the code lines 18 to 30 to prepare the Stata environment.

STEP 1: Declare VAT variables needed for Random Forest (RF) Model

- VAT variables: `rtn_year tax_sales vat_tax_sales zero_rate_sales exempt_sales import_sales vat_import_sales export_sales capital_sales nondeduc_sales taxunregistered_sales vat_taxunregistered_sales total_sales total_vat_sales tax_purchases vat_tax_purchases zero_rate_purchases exempt_purchases import_purchases vat_import_purchases export_purchases capital_purchases nondeduc_purchases taxunregistered_purchases vat_taxunregistered_purchases total_purchases total_vat_purchases vat_payment city region isic sector tax_region`

STEP 2: Declare iterations needed and number of variables to replace for RF Model

- VAT selection: 400 iterations (insert figure in syntax `“iter(400)”`) and 32 variable replacements (insert figure in syntax `“numvar(32)”`)

Set path in line 36 or use similar path as done in Data Prep exercise. In line 47 indicate the data to use as `“vat_audit”`

Exercise 1: Set up and utilities

Aim: How to start and prepare for the gap estimation

1. To set up audit and evasion data for easy identification perform the following cleaning steps:
 - a) Create a dummy variable “reg_audit” for all audits greater than zero.
 - b) Create a new variable “evaded” from the “vat_evasion” variable.
 - c) Run lines 67-69 to create and replace new variables (indicator and mimic_tax).
 - d) Run the RF model using VAT variables and numbers from STEPs 1 and 2.

Exercise 1 – Solutions

- a) Create a dummy variable “reg_audit” for all audits greater than zero.

```
56  
57 generate reg_audit = (audit>0)  
58
```

- b) Create a new variable “evaded” from the “vat_evasion” variable.

```
58  
59 generate evaded = vat_evasion  
60
```

- c) Run lines 67-69 to create and replace new variables (indicator and mimic_tax)

```
67 generate indicator = (evaded !=. & reg_audit==1)  
68 generate mimic_tax = evaded  
69 replace mimic_tax= 0 if evaded==.
```

- d) Run the RF model using VAT variables and numbers from STEPs 1 and 2.

Exercise 1 – Solutions continued

d) Run the RF model using VAT variables and numbers from STEPs 1 and 2.

```
68 rforest mimic_tax year tax_sales vat_tax_sales zero_rate_sales exempt_sales import_sales  
↳ vat_import_sales export_sales capital_sales noneduc_sales taxunregistered_sales  
↳ vat_taxunregistered_sales total_sales total_vat_sales tax_purchases vat_tax_purchases  
↳ zero_rate_purchases exempt_purchases import_purchases vat_import_purchases export_purchases  
↳ capital_purchases noneduc_purchases taxunregistered_purchases vat_taxunregistered_purchases  
↳ total_purchases total_vat_purchases vat_payment city region isic sector tax_region if indicator==1,  
↳ type(reg) iter(400) numvars(32)  
69 predict assessment if indicator==0  
70  
71 replace assessment = evaded if indicator==1 & reg_audit==1
```

Exercise 2 – Estimating VAT gaps by return year

Aim: Run gap estimation by return year.

2. Estimate VAT gaps by return year by editing lines 92-104 based on the following questions:
 - a) Collapse the total assessment (assessment) and vat payment (vat_payment) variables using the return year (rtn_year) variable.
 - b) Generate the potential VAT (v01) by summing the total assessment (assessment) and VAT payment (vat_payment) variables.
 - c) Generate the percentage annual VAT gap (vat_gap01) by dividing the assessment (assessment) by the potential VAT (v01) variable and multiply the result by 100
 - d) Plot the annual percentage vat (vat_gap01) by the return year (rtn_year) and save it in an Excel spreadsheet. Estimate the average annual gaps and export to Excel.

Exercise 2 - Solution

- a) Collapse the total assessment (assessment) and vat payment (vat_payment) variables using the return year (rtn_year) variable.

```
86  
87 collapse (sum) assessment vat_payment, by(rtn_year)
```

- b) Generate the potential VAT (v01) by summing the total assessment (assessment) and VAT payment (vat_payment) variables.

```
88  
89 generate v01 = assessment+abs(vat_payment)
```

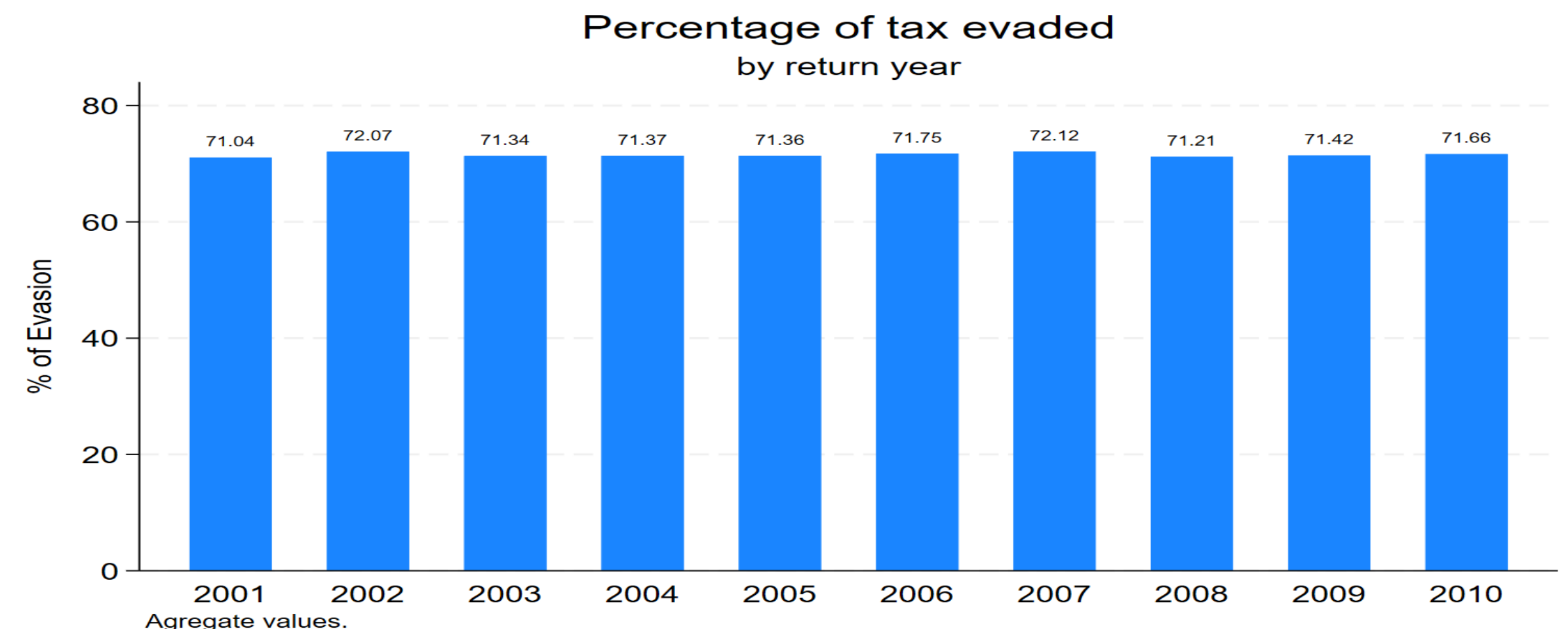
- c) Generate the percentage annual VAT gap (vat_gap01) by dividing the assessment (assessment) by the potential VAT (v01) variable and multiply the result by 100.

```
90  
91 generate vat_gap01 = (assessment/v01)*100
```

Exercise 2 – Solutions continued

- d) Plot the annual percentage vat (vat_gap01) by the return year (rtn_year) and save it in an Excel spreadsheet. Estimate the average annual gaps and export to Excel.

```
94 graph bar (mean) vat_gap01, over(rtn_year) blabel(bar, size(vsmall) format(%9.2f)) legend( label(1 "TAX") ) ytitle(% of Evasion) title("Percentage of tax
95 evaded") subtitle("by return year") note("Agregate values.")
96 graph save Graph "$graphs/vat_gap_year.gph", replace
97 graph export "$graphs/vat_gap_year.pdf", as(pdf) replace
98 estpost tabstat vat_gap01 , by(year) stat(max)col(stat)
99 esttab . using "$xlsout/results_year.csv", cells("max(lab(Max) fmt(%15.2fc))") noobs nonumber eqlabels(`e(labels)') varwidth(20) replace
```



Exercise 3 – Estimating VAT gaps by location (region)

Aim: Run gap estimation by location.

3. Estimate VAT gaps by return year by editing lines 117-129 based on the following questions:
 - a) Collapse the total assessment (assessment) and vat payment (vat_payment) variables using the location (region) variable.
 - b) Generate the potential VAT (v02) by summing the total assessment (assessment) and VAT payment (vat_payment) variables.
 - c) Generate the percentage annual VAT gap (vat_gap02) by dividing the assessment (assessment) by the potential VAT (v02) variable and multiply the result by 100.
 - d) Plot the regional percentage vat (vat_gap02) by the location (region) and save it in an Excel spreadsheet. Estimate the average location gaps and export to Excel.

Exercise 3 – Solutions

- a) Collapse the total assessment (assessment) and vat payment (vat_payment) variables using the location (region) variable.

```
.11  
.12 collapse (sum) assessment vat_payment, by(region)
```

- b) Generate the potential VAT (v02) by summing the total assessment (assessment) and VAT payment (vat_payment) variables.

```
.13  
.14 generate v02 = assessment+abs(vat_payment)
```

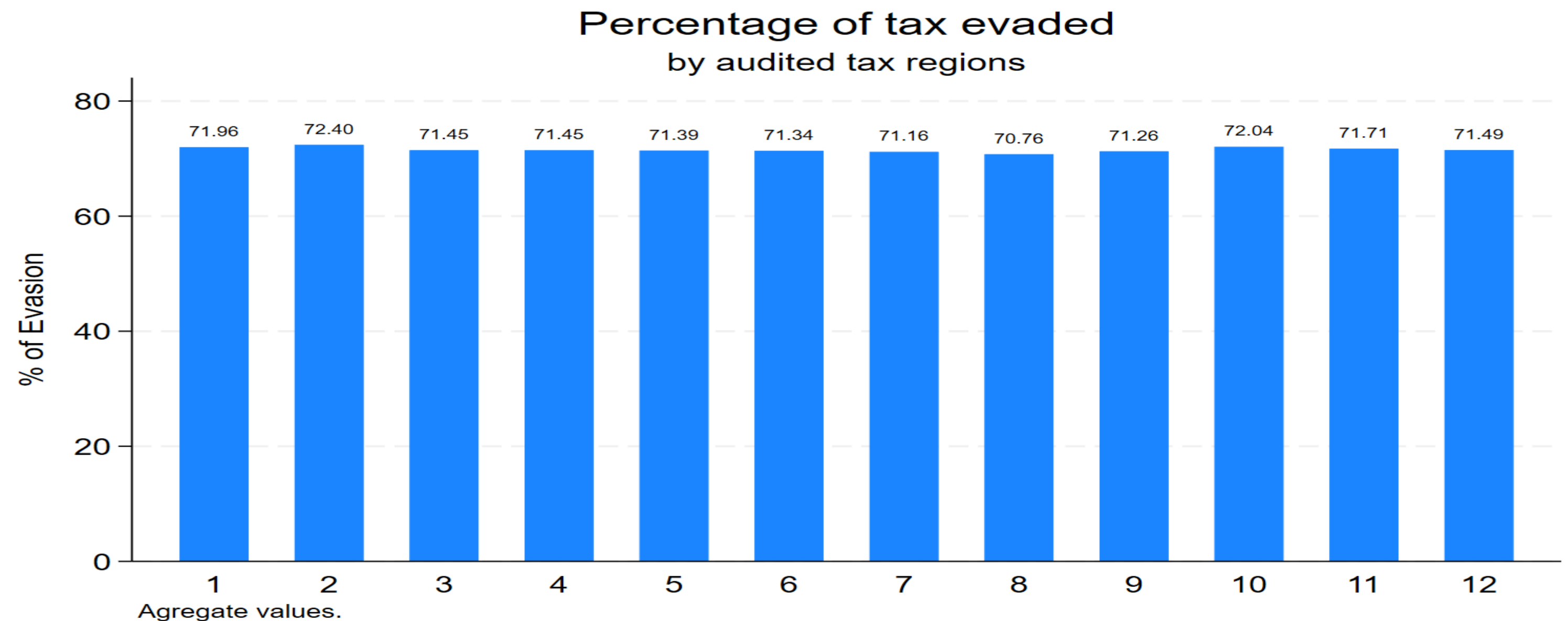
- c) Generate the percentage location VAT gap (vat_gap02) by dividing the assessment (assessment) by the potential VAT (v02) variable and multiply the result by 100.

```
.15  
.16 generate vat_gap02 = (assessment/v02)*100
```

Exercise 3 – Solutions continued

- d) Plot the annual percentage vat (vat_gap02) by the location (region) and save it in an Excel spreadsheet. Estimate the average location gaps and export to Excel.

```
118
119 graph bar (mean) vat_gap02, over(region) blabel(bar, size(vsmall) format(%9.2f)) legend( label(1 "TAX") ) ytitle(% of Evasion) title("Percentage of tax
    evaded") subtitle("by audited tax regions") note("Agregate values.")
120 graph save Graph "$graphs/vat_gap_loc.gph", replace
121 graph export "$graphs/vat_gap_loc.pdf", as(pdf) replace
122
123 estpost tabstat vat_gap02 , by(region) stat(max)col(stat)
124 esttab . using "$xlsout/results_loc.csv", cells("max(lab(Max) fmt(%15.2fc))") noobs nonumber eqlabels(`e(labels)') varwidth(20) append
```



Exercise 4 – Potential VAT with Plots

Aim: Visualizing tax gap effects and consequences

4. Generate potential VAT revenues and plots by return year and location.
This allows us to quantify the potential revenue and the erosion by tax evasion.

a)

1. Calculate annual VAT (assessment + actual VAT payments)
2. Plot it with total assessment assessments and actual VAT payments by return year.
3. Estimate the averages for potential VAT VAT assessments and actual VAT by return year and export to Excel.

HINT: First Collapse the total assessment (assessment) and vat payment (vat_payment) variables using the return year (rtn_year) variable.

Exercise 4a – Solutions

- a) HINT: First Collapse the total assessment (assessment) and vat payment (vat_payment) variables using the return year (rtn_year) variable.

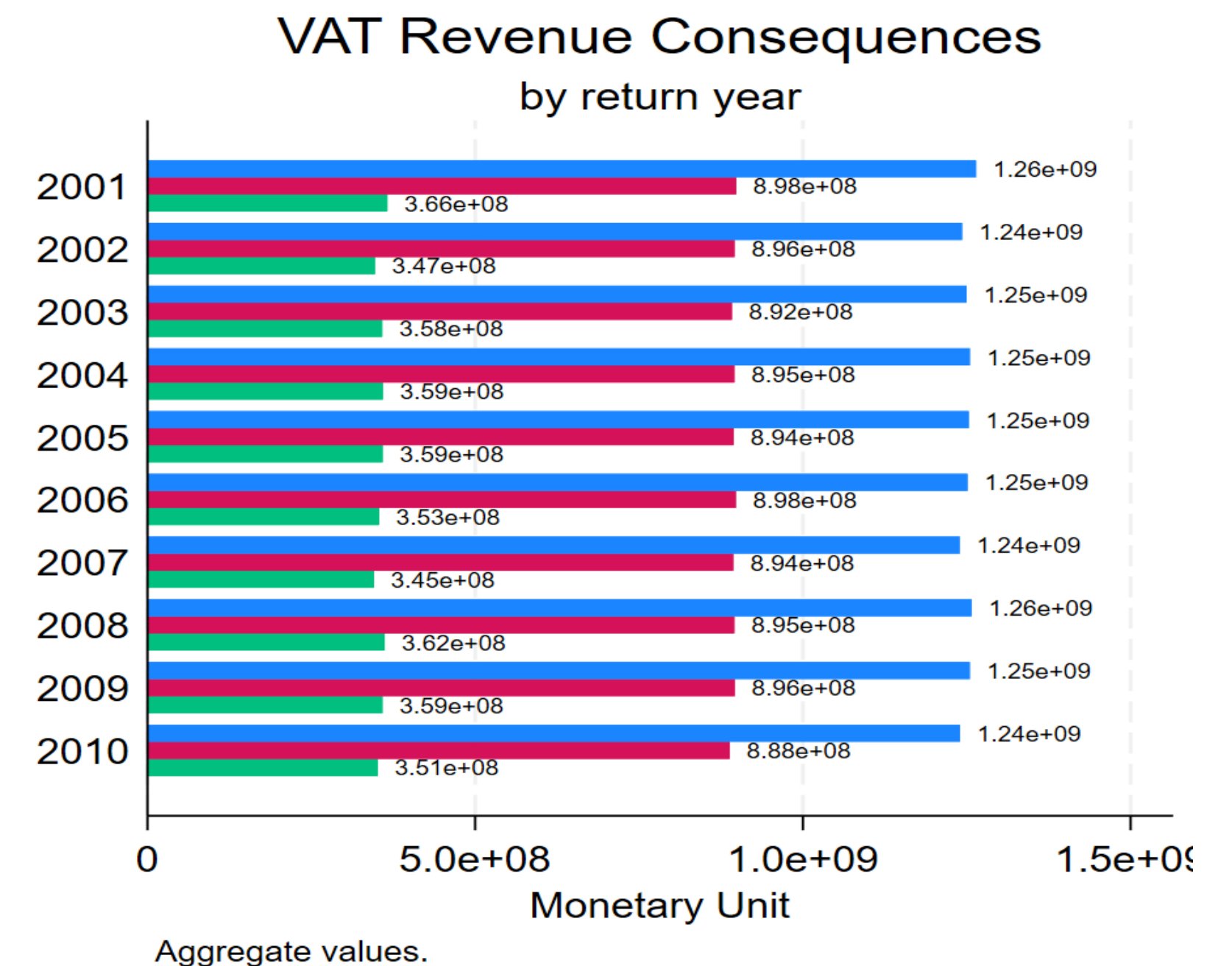
```
136 collapse (sum) assessment vat_payment, by(rtn_year)
```

1. Calculate total annual potential VAT (assessment + actual VAT payments)

```
138 generate potential_revenue = assessment + vat_payment
```

2. Plot it with total assessment assessments and actual VAT payments by return year.

```
141 graph hbar (mean) potential_revenue assessment vat_payment, over(rtn_year) blabel(bar, size(
↳ vsmall) format(%9.2f)) legend( label(1 "VAT Potential Revenue") label(2 "VAT Revenue Lost
↳ (predicted)") label(3 "VAT Revenue Collected")) ytitle(Monetary Unit) title("VAT Revenue
↳ Consequences") subtitle("by return year") note("Aggregate values.")
142 graph save Graph "$graphs/revenue_gap_year.gph", replace
143 graph export "$graphs/revenue_gap_year.pdf", as(pdf) replace
```



Exercise 4a – Solutions continued

a) Generate potential VAT revenues and plots by return year and location. This allows us to quantify the potential revenue and the erosion by tax evasion.

3. Estimate the averages for potential VAT VAT assessments and actual VAT by return year and export to Excel.

```
145 estpost tabstat potential_revenue assessment vat_payment , by(rtn_year) stat(max)col(stat)
146 esttab . using "$xlsout/results_nominal_year.csv", cells("max(lab(Max) fmt(%15.2fc))") noobs
    ↪ nonumber eqlabels(`e(labels)') varwidth(20) replace
```

Exercise 4 – Potential VAT with Plots

Aim: Visualizing tax gap effects and consequences

4. Generate potential VAT revenues and plots by return year and location. This allows us to quantify the potential revenue and the erosion by tax evasion.

b)

1. Calculate regional VAT (assessment + actual VAT payments).
2. Plot it with total assessment assessments and actual VAT payments by location (region).
3. Estimate the averages for potential VAT VAT assessments and actual VAT by location and export to Excel.

HINT: First Collapse the total assessment (assessment) and vat payment (vat_payment) variables using the location (region) variable.

Exercise 4b – Solutions

- a) HINT: First Collapse the total assessment (assessment) and vat payment (vat_payment) variables using the location (region) variable.

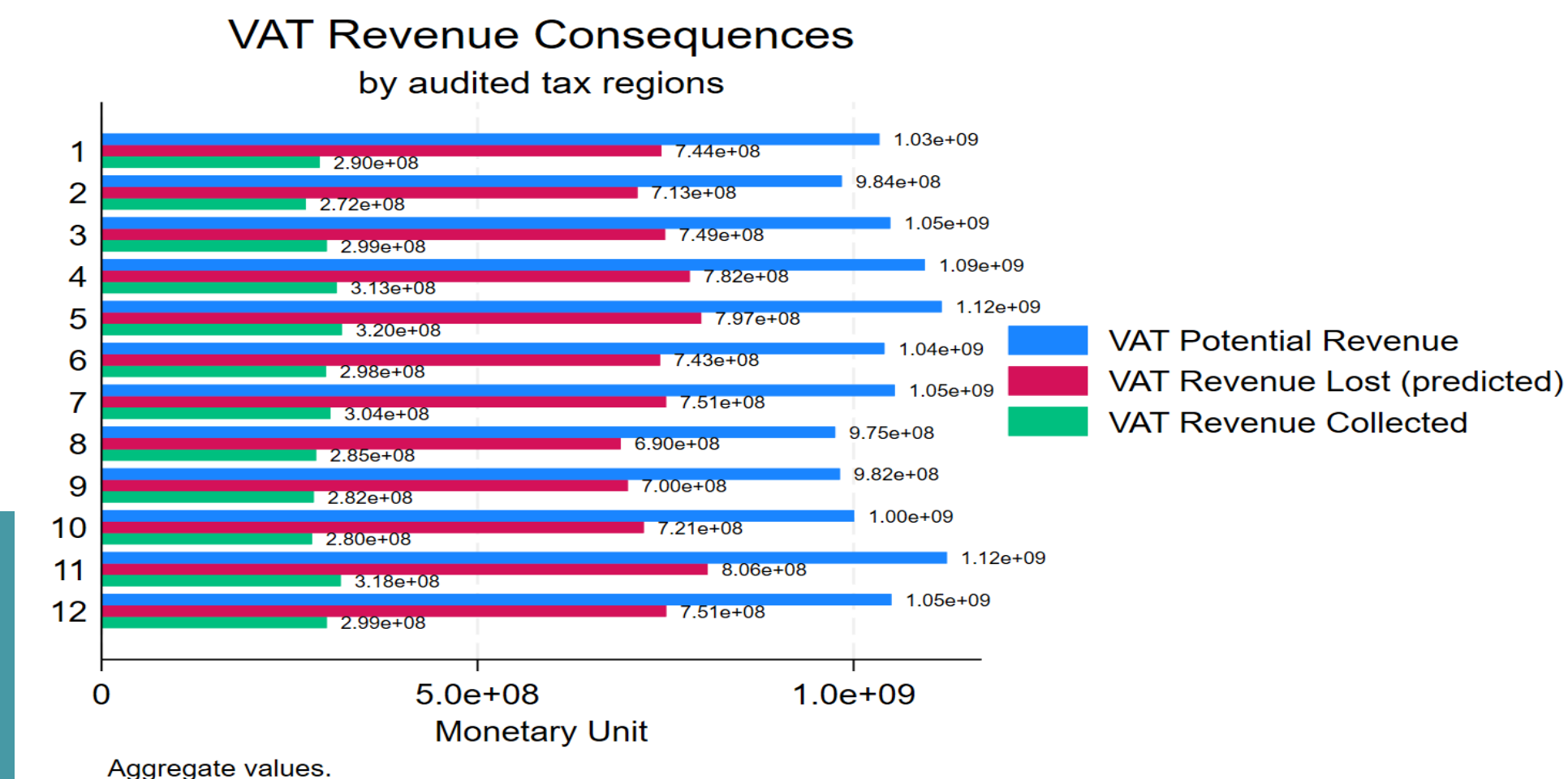
```
155 collapse (sum) assessment vat_payment, by(region)
```

1. Calculate regional VAT (assessment + actual VAT payments)

```
157 generate potential_revenue = assessment + vat_payment
```

2. Plot it with total assessment assessments and actual VAT payments by return year.

```
160 graph hbar (mean) potential_revenue assessment vat_payment, over(region) blabel(bar, size(
↳ vsmall) format(%9.2f)) legend( label(1 "VAT Potential Revenue") label(2 "VAT Revenue Lost
↳ (predicted)") label(3 "VAT Revenue Collected") ) ytitle(Monetary Unit) title("VAT Revenue
↳ Consequences") subtitle("by audited tax regions") note("Aggregate values.")
161 graph save Graph "$graphs/revenue_gap_loc.gph", replace
162 graph export "$graphs/revenue_gap_loc.pdf", as(pdf) replace
```



Exercise 4b – Solutions continued

a) Generate potential VAT revenues and plots by return year and location. This allows us to quantify the potential revenue and the erosion by tax evasion.

3. Estimate the averages for potential VAT VAT assessments and actual VAT by location and export to Excel.

```
164 estpost tabstat potential_revenue assessment vat_payment , by(region) stat(max)col(stat)
165 esttab . using "$xlsout/results_nominal_loc.csv", cells("max(lab(Max) fmt(%15.2fc))") noobs
    ↪ nonumber eqlabels(`e(labels)') varwidth(20) append
```

Homework: Estimating CIT and PAYE Gaps

Aim: Test understanding for other tax head gap estimations.

Repeat exercises using CIT and PAYE datasets and variable list and iterations below:

CIT variables: rtn_year fixed_assets depreciation investment loans provisions dividends reserves debts sales_turnover wages total_purchases profits operating_cost taxes_paid donations total_profit taxable_profits cit_payment city region isic sector tax_region

CIT selection: 400 iterations (insert figure in syntax “iter(400)”) and 20 variable replacements (insert figure in syntax “numvar(20)”)

PAYE variables: rtn_year number_employers total_gross_payment total_nontax_payment total_tax_deductions total_payroll_tax total_chargeable_payments paye_tax paye_payment city region isic sector.

PAYE selection: 400 iterations (insert figure in syntax “iter(400)”) and 6 variable replacements (insert figure in syntax “numvar(6)”)

Thank you!

Kwabena Adu-Ababio

kwabena@wider.unu.edu



**UNU
WIDER**

More information available online

wider.unu.edu

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