Measuring BEPS: MNEs vs. comparable non-MNEs method

Italian case study - Practice

Federico Sallusti
Istat

Building up the dataset

- The database for the analysis is composed of three informative sources:
  - The archive **Frame-SBS**, which includes the information about the structure and economic variables for the whole set of 4.4 millions of firms
  - The archive **COE-TEC** (Integrated International Trade Database), which includes the information about imports and exports (by product and origin/destination country) for the whole set of firms
  - The archive **ASIA-Groups** (Italian version of European EGR), which includes the information about firms involved in domestic and foreign groups
- For each unit in the business system, the final database reports comprehensive information about:
  - The economic and organizational structure
  - The characteristics of its inclusion in the network of international trade
  - If applies, the positioning within MNE groups

Building up the dataset

- Frame-SBS contains about 4.3 million units for 2019
- COE-TEC cointains about 4.3 million units for 2019 (165515 are internationalised, i.e. positive values of imports and/or exports)
- Asia-Groups contains 233092 units
- Some lilmitation is applied to the dataset (i.e. exclusion of units with 0 workers, value added lower than 0, missing relevant variables, sectors with peculiar characteristics such as tobacco, financial auxiliaries, coke and refineries)
- The final dataset contains 3829908 observation

Building up the dataset

Unit ID	NACE	Workers	Size class	Value added	Turn-over	Salaries	Intermediate costs	Costs for goods	Costs for services	Costs for royalties	Costs for R&D	Costs from sub-contracting	Revenues from sub-contracting	EBIT	Value of imports	Value of exports	Group ID	Type of group	Nationality of headquarter	Nationality of units
xxx	kk291	8.9	2	374.3	619.1	456.7	244.8	14.1	223.1	0.2	0.0	0.8	0.1	-82.4	0.0	1.8	.	.	.	.
xxx	kk220	22.1	4	1634.5	2281.5	1513.2	647.0	319.4	207.0	0.1	0.1	71.1	0.0	121.3	0.0	0.0	.	.	.	.
xxx	kk220	25.4	4	1775.9	2292.3	1902.0	516.4	0.0	263.7	0.1	0.2	90.6	0.0	-126.2	0.0	0.0	.	.	.	.
xxx	kk292	10.6	3	980.4	4419.4	638.3	3439.0	1.9	3288.8	2.4	0.0	114.4	14.3	342.2	0.0	16.8	.	.	.	.
xxx	kk215	1.0	1	29.8	77.7	15.1	47.8	0.0	24.1	0.0	0.0	0.3	0.0	14.7	0.0	0.0	.	.	.	.
xxx	kk291	8.7	2	523.1	2931.1	387.0	2408.0	5.6	2291.2	1.8	0.4	8.6	0.7	136.1	1.1	58.2	.	.	.	.
xxx	kk292	13.7	3	575.4	12121.5	485.4	11546.1	29.4	11361.9	8.5	0.0	395.2	38.6	90.0	0.0	0.0	.	.	.	.
xxx	kk292	3.0	1	112.1	855.9	95.9	743.7	0.7	723.0	0.3	0.0	4.7	0.5	16.2	0.0	9.7	.	.	.	.
xxx	kk230	8.0	2	181.7	935.4	271.2	753.7	230.3	418.9	0.3	0.1	1.6	0.2	-89.5	2.2	7.5	.	.	.	.
xxx	kk219	6.8	2	248.7	433.7	156.7	185.0	37.1	123.9	0.1	0.0	0.5	0.1	91.9	0.0	0.0	.	.	.	.
xxx	kk215	1.9	1	18.0	75.0	1.0	57.0	19.7	22.1	0.0	0.0	0.1	0.0	17.0	0.0	0.0	.	.	.	.
xxx	kk292	323.1	7	6499.0	7044.0	11550.0	563.0	0.0	508.0	0.0	0.0	0.0	0.0	-5051.0	0.0	0.0	.	.	.	.
xxx	kk215	8.5	2	1295.0	2306.7	356.7	710.7	13.8	585.8	0.5	0.1	2.2	0.4	938.3	0.0	0.0	zzz	MULES	BE	IT
xxx	kk292	33.8	4	1626.4	8169.8	1623.5	6543.4	54.8	5163.7	2.1	3.2	1774.0	0.0	2.9	0.0	0.0	zzz	MULES	DE	IT
xxx	kk292	4.1	1	0.0	1274.6	177.5	1383.4	11.7	1205.0	0.1	0.0	5.0	0.5	-177.5	0.0	0.0	zzz	MULES	DE	IT
xxx	kk219	6.0	2	461.4	659.2	355.5	197.9	3.8	134.6	0.1	0.0	0.5	0.1	105.9	0.0	0.0	zzz	MULES	NL	IT
xxx	kk101	19.7	3	3543.2	11669.4	1133.2	8126.1	392.4	6772.4	3.4	4.1	477.7	103.6	2410.0	0.0	0.0	zzz	MULIT	IT	IT
xxx	kk292	37.3	4	9024.1	32235.9	2442.0	23981.4	2054.4	20076.2	8.3	12.4	6897.0	0.0	6582.1	4482.6	5.7	zzz	MULIT	IT	IT
xxx	kk212	539.3	7	390475.0	531493.0	37243.0	269003.0	2549.0	85701.0	0.0	0.0	0.0	0.0	353232.0	0.0	0.0	zzz	MULIT	IT	IT
xxx	kk230	237.2	6	38712.6	67133.1	13927.6	28420.5	1211.9	22238.3	0.0	0.0	0.0	0.0	24785.0	0.0	0.0	zzz	MULIT	IT	IT
xxx	kk292	4.7	1	82.6	260.6	189.3	178.0	2.4	149.0	0.0	0.0	0.6	0.1	-106.7	0.0	0.0	zzz	MULIT	IT	IT
xxx	kk291	3.5	1	105.7	571.3	105.7	465.6	0.0	421.5	0.0	0.0	1.8	0.2	-0.1	0.0	0.0	zzz	MULIT	IT	IT
xxx	kk291	82.4	5	3029.8	27788.8	3552.9	24889.4	59.0	24150.2	5.9	28.4	2524.1	28.0	-523.1	0.0	49.3	zzz	MULIT	IT	IT
xxx	kk220	7.2	2	1374.7	3981.1	429.3	2606.4	1770.2	621.2	0.5	0.1	2.3	0.8	945.4	0.0	0.0	zzz	MULIT	IT	IT
xxx	kk292	369.8	7	32095.0	59253.0	19824.0	30490.0	312.0	24608.0	0.0	0.0	0.0	0.0	12271.0	7.0	4.5	zzz	MULIT	IT	IT

Frame-SBS | COE-TEC | Asia-Groups

See Tables **DB exemple** and **Descriptive**

Overview of the method by step

- MNE vs. comparable non-MNE method is composed by three phases:

  1. The **identification** of the either BEPS generating (outward IFFs) or BEPS collecting (inward IFFs) nature of the country (OECD's dashboard approach of BEPS indicators)

  2. The **selection** of tax avoiding (TA) units among MNEs

     - Italian MNEs are evaluated in order to define if they are suspected of tax avoiding behaviour based on the comparison between MNEs and a control group consisting of (comparable) non-MNEs

  3. The **correction** of profits for TA MNEs

     - The EBIT-to-turnover ratio of TA units is adjusted exploting the selection model in order to compare the economic results of TA MNEs vs. the one of non-TA MNEs

Selection - Overview

- The phase of selection is composed of three steps:
  - **Control group definition**

    For each MNE unit, a control group of domestic firms is defined using propensity score matching

  - **Between comparison (MNEs vs. non-MNEs)**

    For each pair MNE unit-control group, a comparison in terms of profit share is used to define a proxy variable, which stresses possible abnormal behaviours by MNEs

  - **Within comparison (among MNEs)**

    ROC analysis is used to define the final clustering between tax avoiding (TA) and non tax-avoiding (NTA) units starting from the proxy variable

Selection - Definition of control groups

Definition of confounding variables for PS matching analysis

- **v1** = Turnover / Workers
- **v2** = Number of workers
- **v3** = Costs for goods / Total intermediate costs
- **v4** = Value of exports / Turnover
- **v5** = Value of imports / Total costs
- **v6** = Salaries / (Salaries + Total intermediate costs)
- **v7** = Costs for services / Total intermediate costs
- **vv** = EBIT / Turnover

See Table **DB PS indicators**

Selection - Definition of control groups

Propensity score matching analysis

- PS Model to define matching probabilities

treat(treated='1') = v1 v2 v3 v4 v5 v6 v7 NUTS2 (Logit model to define matching probabilities)

- Matching method and number of similar

match method = greedy (k = **5**)

- Binding characteristics

Exact (NACE3 size class NUTS2)

See Table **Outmatch**

Selection - Between comparison

Definition of the proxy of suspect

- Prospensity score matching allows to define a control group of domestic firms for each MNE unit
- For each pair MNE unit-control group, a **proxy of suspect** of TA is given by the following condition:
  - **Proxy = 1**

    if ebit-to-turnover ratio for the MNE unit is lower than the average of the control group

  - **Proxy = 0**

    if ebit-to-turnover ratio for the MNE unit is greater or equal to the average of the control group

See Table **Proxy**

Selection - Within comparison

Steps in within comparison

- Definition of variables

From structural and performance characteristics of MNEs it selects the variables of interests in capturing the behavior of MNEs

- Factor Analysis

From x0-x8 variables it selects two factors

- Definition of composite indicator

From factors it defines the composite to be used in the logit model of the ROC analysis

- ROC analysis

It allows to define the final classification between Tax Avoiding (TA) and non-Tax Avoiding (nonTA) MNEs

Selection - Within comparison

Definition of variables (inverse relationships with proxy, the higher the values the lower the probability of TA)

- **x0** = EBIT / Turnover
- **x1** = Value added / Turnover
- **x2** = Costs on R&D / Total intermediate costs
- **x3** = 1 – (Costs on royalties / Turnover)
- **x4** = 1 – Value of imports / Total intermediate costs
- **x5** = 1 – Tax framework (by Country, differential)
- **x6** = Salaries / Turnover
- **x7** = 1 – Cost for services / Turnover
- **x8** = Value of exports / Turnover

See Table **DB ROC indicators**

Selection - Within comparison

Factor analysis and the definition of the composite indicator

- **x0 - x8** variables are firstly standardized in order to avoid scale effects
- From factor analysis, the first two auto-rotated factors are retained and the composite is defined as follows

$$I_i = \omega_1 \left(\sum_j \gamma_{j,1} x_{j,i} \right) + \omega_2 \left(\sum_j \gamma_{j,2} x_{j,i} \right)$$

where, $\gamma_{j,1}$ and $\gamma_{j,2}$ are the loadings of the j-th variable in factors 1 and 2, $x_{j,i}$ is the value of the j-th variable for the i-th observation, and ω_1 and ω_2 are weights in term of explained variance

	Factor1	Factor2
x0	0.043	0.518
x1	0.057	0.519
x2	-0.086	0.025
x3	0.452	-0.063
x4	-0.043	0.178
x5	-0.011	0.050
x6	0.195	-0.002
x7	0.465	-0.049
x8	-0.063	0.118
Explained variance	2.098	1.688

Selection - Within comparison

ROC analysis

- Using the proxy of suspect and the composite indicator, the following logit model can be run:

$$Prob\ (Proxy = 1|C)_i = \Lambda(\alpha C)_i$$

Observations			1124
Response profile			
Proxy	0	Frequency	237
	1		887
Criteria			
		Intercept only	Intercept and covariates
AIC		1159.9	895.9
SC		1164.9	905.9
-2 Log L		1157.9	891.9
Results			
Parameters	Estimate	Standard error	P-value
Composite	-2.4225	0.2075	<.0001
Goodness of fit			
Percentuale concordi	82.4	D di Somers	0.648
Percentuale discordi	17.6	Gamma	0.648

Selection - Within comparison

- The ROC analysis can be traced back to classification problems in which classifiers can give the four possible outcomes shown in the **confusion matrix.** The efficiency of the classifier can be measured using two metrics:
- **Sensitivity** measures the ability of the classifier to detect true positives, i.e. $TP/(TP + FN)$;
- **Specificity** measures the ability of the classifier to detect true negatives, i.e. $TN/(TN + FP)$, where it is usually considered in its reciprocal expression (1-Specificity), which measures the correct detection of false positives

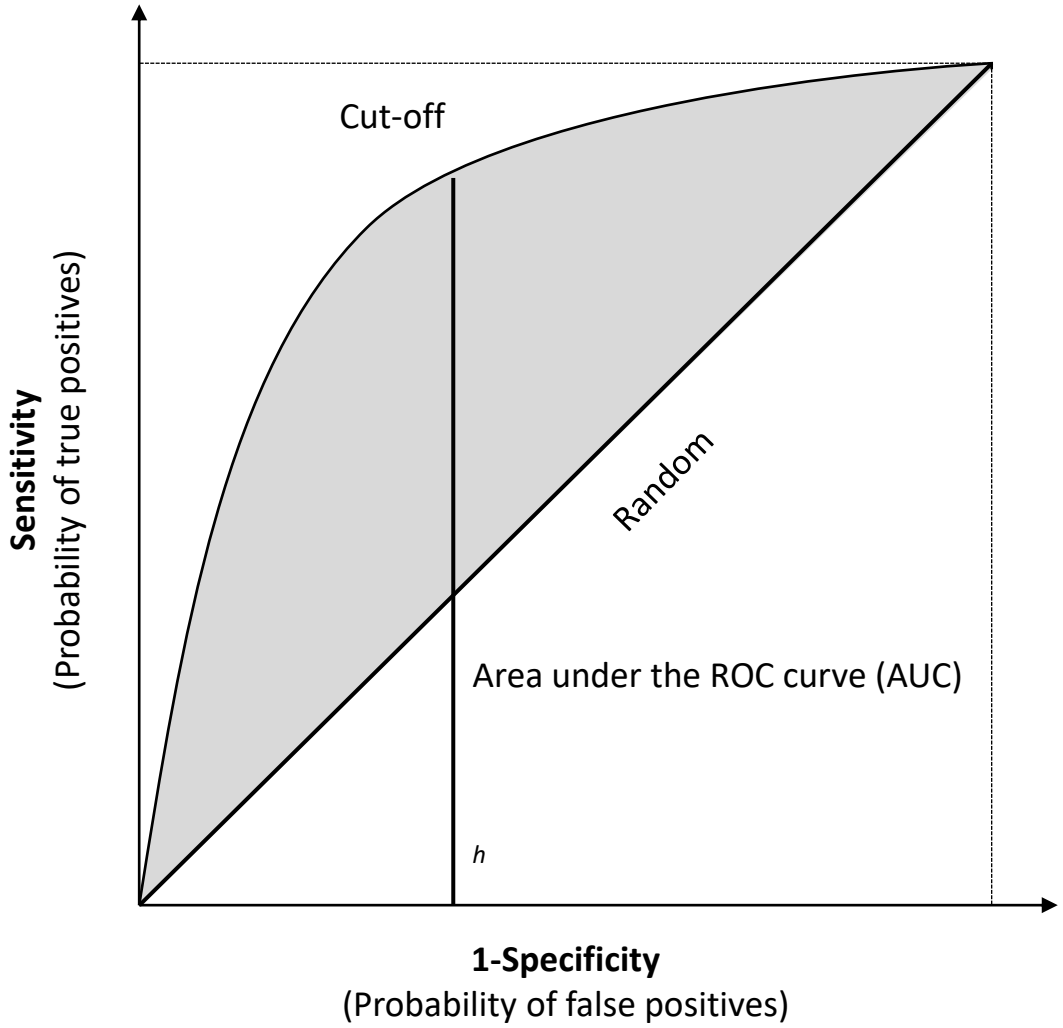
		Estimated classification	
		1	0
True classification	1	*TP*	*FN*
	0	*FP*	*TN*

Selection - Within comparison

- Considering a logit model having:
  - a binomial dependent reflecting a given status
  - a classifier represented by a single (even composite) indicator

the distribution of probabilities resulting from the logit estimates can be displayed in the space of Sensitivity and 1-Specificity by the **ROC curve**

- The line of the ROC curve represents the probabilities assigned by the model to each observation in the space of the trade-off between the probability of detecting true or false positives across all possible cut-off points along the values of the classifier

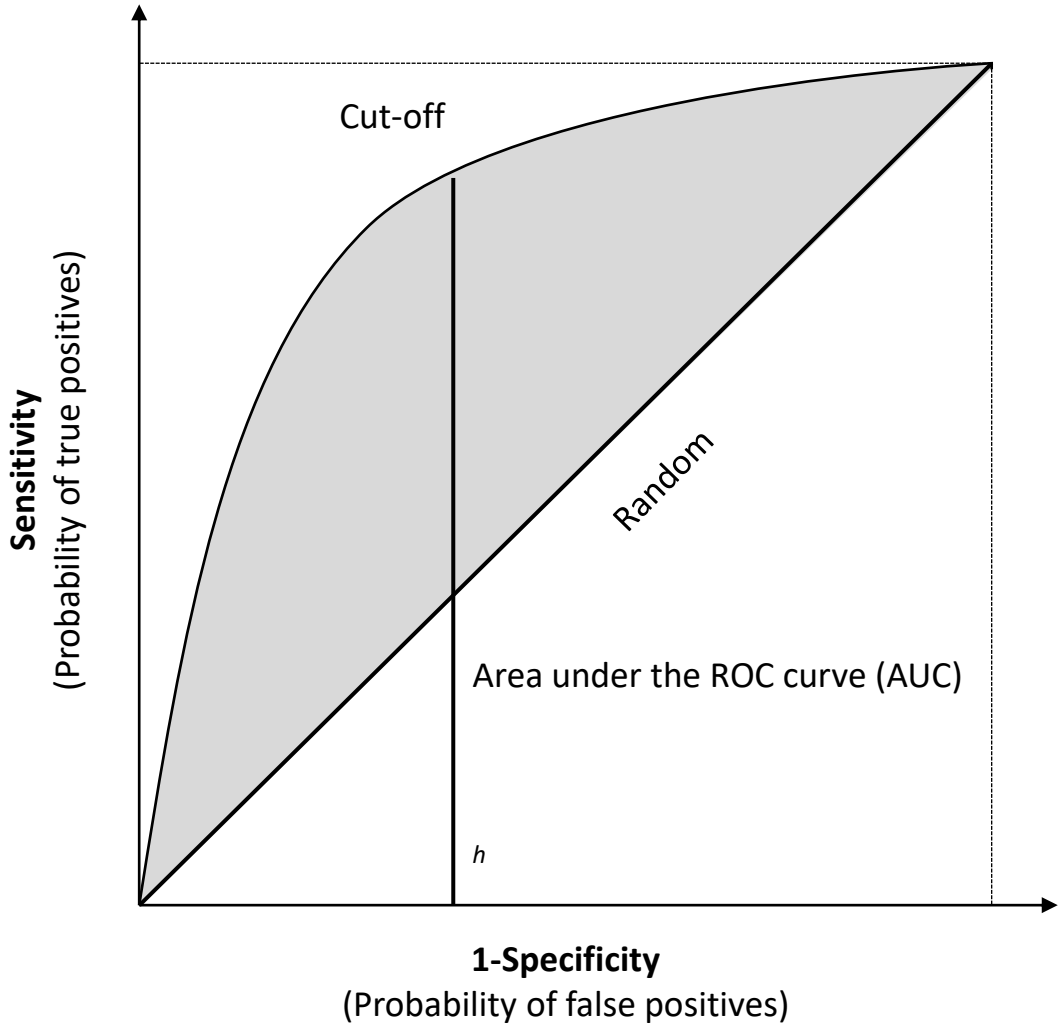


Selection - Within comparison

- In order to single out, along the ROC curve, the observation that most efficiently discriminates between positives and negatives (Cut), the following equation should be maximized:

$$Cut = h * sensitivity - (1 - h) * (1 - specificity)$$

where h and (1-h) represent the relative weights to manage the trade-off between true and false positives.



Selection - Within comparison

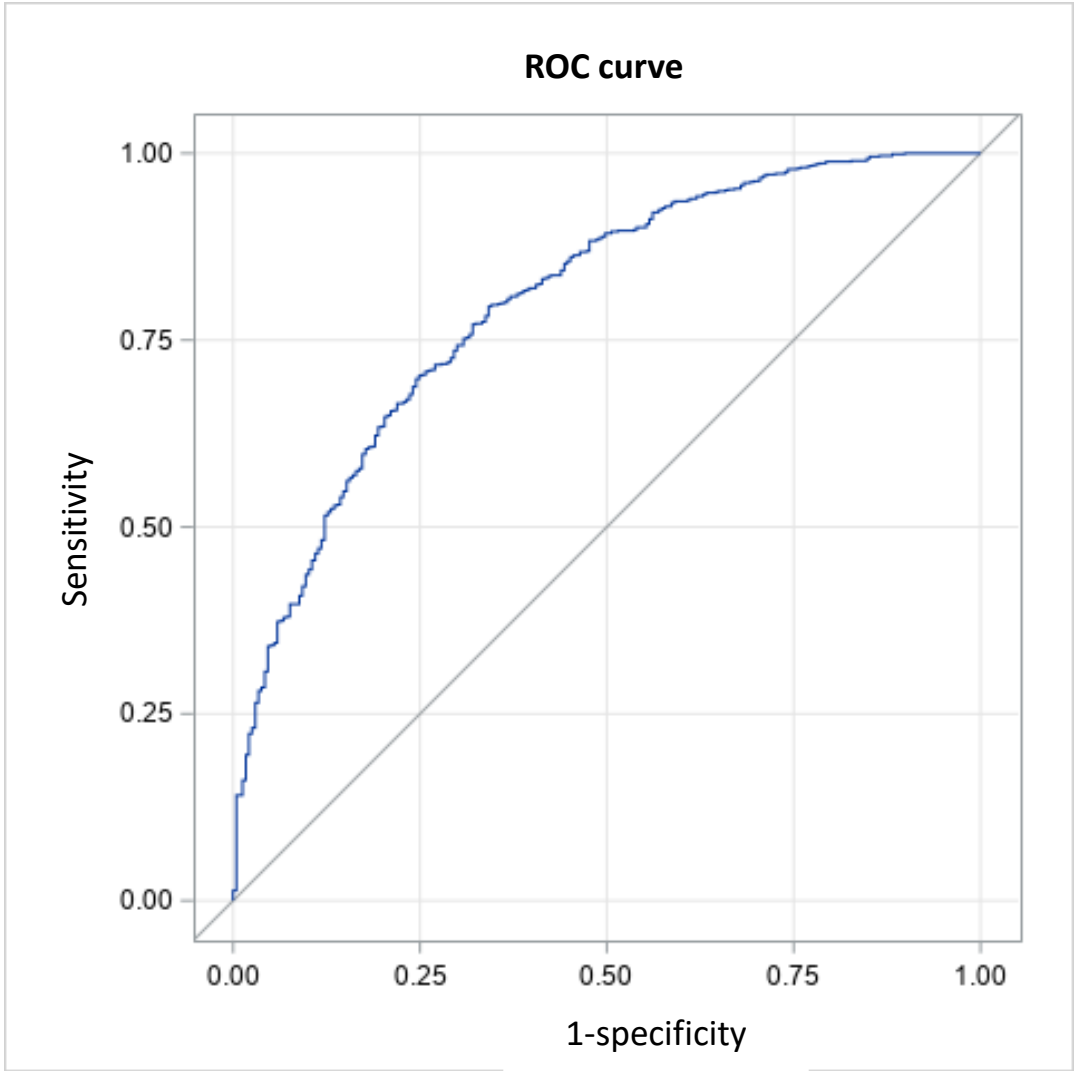
ROC analysis

- The logit model generates the following ROC curve (with AUC=0.8119)

$$Cut = h * sensitivity - (1 - h) * (1 - specificity)$$

with $h = 0.5$ (neutral selection), the ID of the threshold observation can be obtained:

Threshold	ID	Value of the composite
1	656	0.058488



See Tables **AUCs** and **Coeffs and thresholds**

Selection - Within comparison

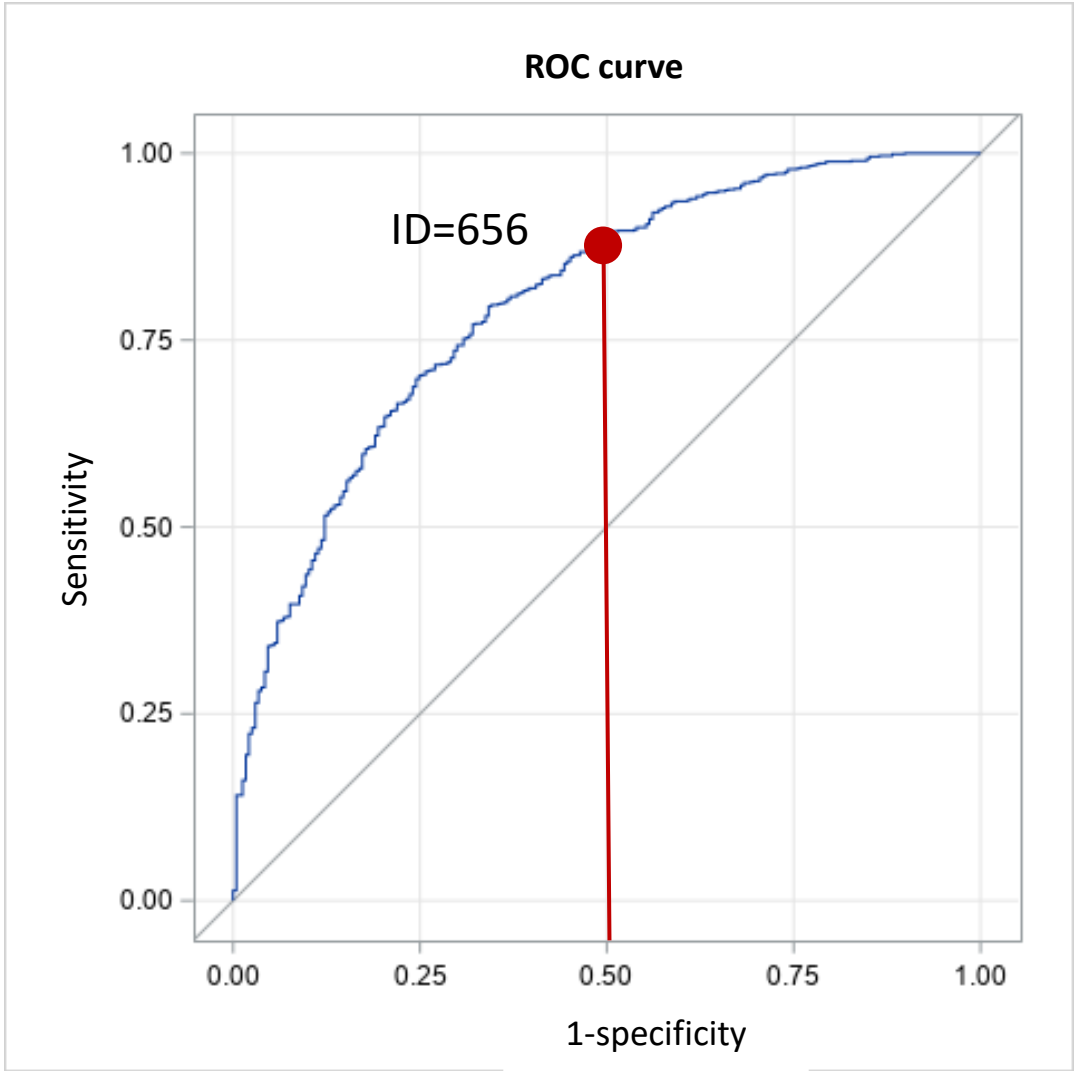
Classification of MNEs

- The value of the composite of threshold MNE unit $\bar{S}$ can be used to classify other observation.

In particular:

- If $I_i \geq \bar{S}$ then the MNE unit is non-TA
- If $I_i < \bar{S}$ then the MNE unit is TA

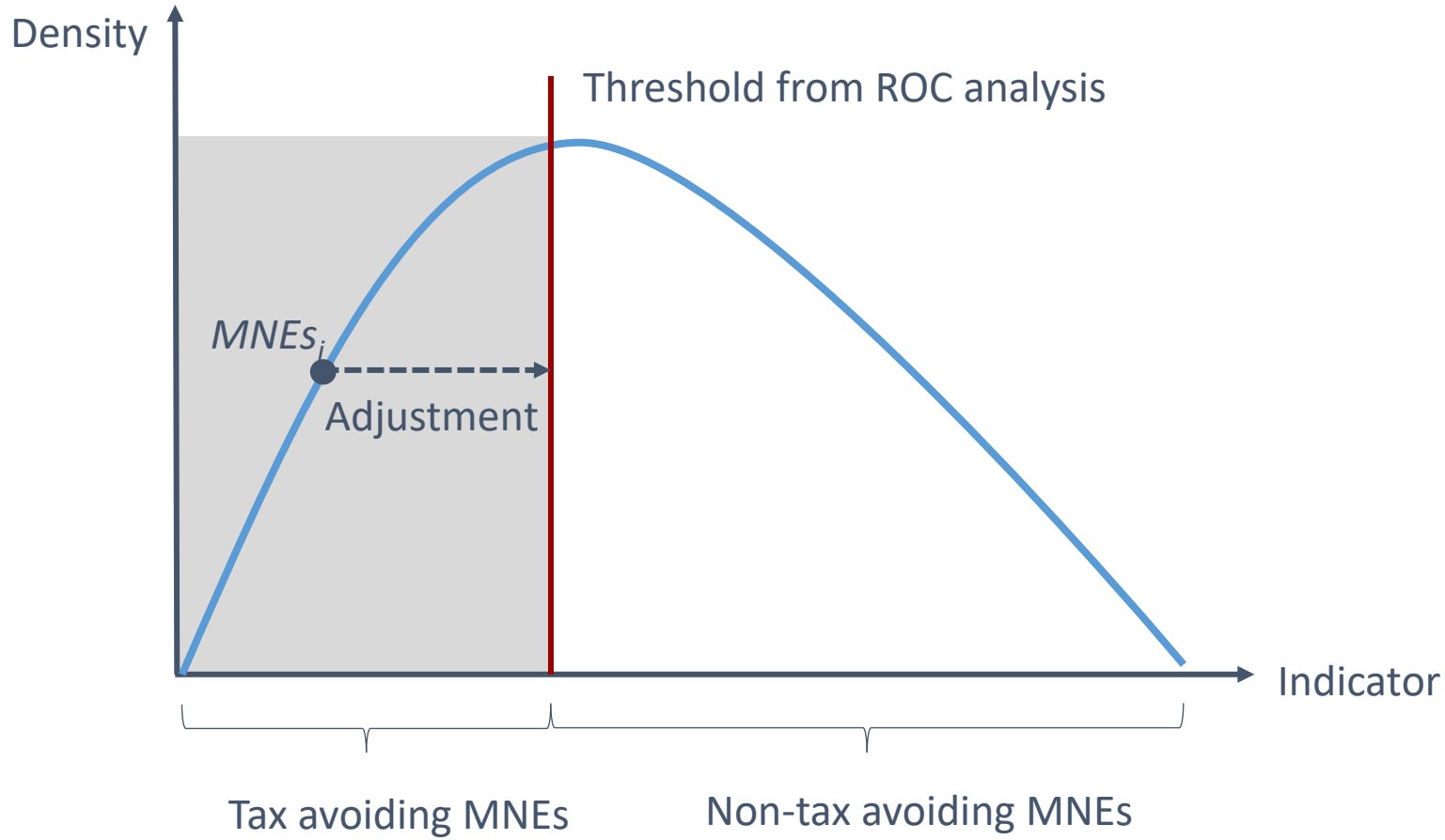
Status		Frequency	Percentage	Cumulative frequency	Cumulative percentage
0	Non-TA	441	39.2	441	39.2
1	TA	683	60.8	1124	100.0



See Table **Fitting**

Correction – conceptual framework

- The correction for BEPS exploits the information provided by the ROC analysis in the selection phase



Correction – conceptual framework

- The correction of profit shifting exploits the information provided by the ROC analysis in the selection phase
- For each TA unit, the following condition applies:

$$\bar{S} > \alpha F_{1,i} + \beta F_{2,i}$$

where factors are:

$$F_{1,i} = \sum_j \gamma_{j,1} x_{j,i} \qquad \text{and} \qquad F_{2,i} = \sum_j \gamma_{j,2} x_{j,i}$$

- The procedure assigns to the indicator x_1, which is the ebit-to-turnover ratio, the value such that, for each TA MNEs, the follwing condition is obtained:

$$\bar{S} = \alpha F_{1,i} + \beta F_{2,i}$$

Correction – conceptual framework

- This allows to define the adjustment condition as:

$$\tilde{x}_{j,i} = \frac{\bar{s} - (\alpha \sum_{-j} \gamma_{-j,1} x_{-j,1} + \beta \sum_{-j} \gamma_{-j,2} x_{-j,2})}{\alpha \gamma_{j,1} + \beta \gamma_{j,2}}$$

where:

- $\bar{S}$ is the threshold value defined by the ROC analysis on the composite indicator
- $(\alpha \sum_{-j} \gamma_{-j,1} x_{-j,1} + \beta \sum_{-j} \gamma_{-j,2} x_{-j,2})$ represents the effect of the other variables on the value of the composite indicator
- $\alpha \gamma_{j,1} + \beta \gamma_{j,2}$ represents the weight of the ebit-to-turnover ratio on the value of the composite indicator
- $\tilde{x}_{j,i}$ is the adjusted value of the ebit-to-turnover ratio in order to bring the TA MNE on the threshold
- The amount of the adjustment is obtained as: $(\tilde{x}_{j,i} - x_{j,i}) * Turnover_{\ i}$

Measuring (outward and inward) IFFs

- The amount of the correction is obtained by comparing the EBIT-to-turnover ratio of the two groups of MNEs defined by the model
- The amount of correction actually represents the measure of IFFs
- In particular:

BEPS generating country

$$OutwardIFFs_i = \left(\tilde{x}_{h,i} - x_{j,i}\right) * Turnover_i$$

where $\tilde{x}_{h,i} > x_{j,i}$

BEPS collecting country

$$InwardIFFs_i = -\left(\tilde{x}_{h,i} - x_{j,i}\right) * Turnover_i$$

where $\tilde{x}_{h,i} < x_{j,i}$

See Table **Results, Countries** and **Descriptives**

Thank you.