

# Mesurer l'érosion de la base d'imposition et le transfert de bénéfices (BEPS), Méthode des EMN vs. non-EMN comparables

*Étude de cas italienne - Pratique*

Federico Sallusti

Istat

# Constitution du jeu de données

- La base de données pour l'analyse est composée de trois sources de données :
  - L'archive **Frame-SBS**, qui comprend les informations sur la structure et les variables économiques pour l'ensemble des 4,4 millions d'entreprises.
  - L'archive **COE-TEC** (Integrated International Trade Database), qui comprend les informations sur les importations et les exportations (par produit et, pays d'origine/de destination) pour l'ensemble des entreprises
  - L'archive **ASIA-Groups** (version italienne de l'EGR européen (répertoire EuroGroups), qui comprend des informations sur les entreprises impliquées dans des groupes nationaux et étrangers
- Pour chaque unité du système d'entreprise, la base de données finale fournit des informations complètes sur:
  - La structure économique et organisationnelle
  - Les caractéristiques de son inclusion dans le réseau du commerce international
  - Le cas échéant, le positionnement au sein des groupes d'entreprises multinationales

# Constitution du jeu de données

- L'archive Frame-SBS La SSE renferme environ 4,3 millions d'unités pour 2019.
- L'archive COE-TEC renferme environ 4,3 millions d'unités pour 2019 (165515 sont internationalisées, c'est-à-dire que les valeurs des importations et/ou des exportations sont positives).
- L'archive Asia-Groups comprend 233092 unités
- Certaines restrictions sont appliquées à l'ensemble des données (par exemple, l'exclusion des unités ayant 0 travailleur, une valeur ajoutée inférieure à 0, des variables pertinentes manquantes, des secteurs ayant des caractéristiques particulières comme le tabac, les auxiliaires financiers, le coke et les raffineries).
- L'ensemble final de données comprend 3829908 observations

# Constitution du jeu de données

ID de l'unité	NACE	Employés	Catégorie de taille	Valeur ajoutée	Chiffre d'affaires	Salaires	Frais intermédiaires	Coût des marchandises	Frais de services	Frais de redevances	Coûts pour la R&D	Frais de sous-traitance	Recettes de la sous-traitance	BAIL	Valeur des importations	Valeur des exportations	Identifiant de groupe	Type de groupe	Nationalité du siège social	Nationalité des unités
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	...	...	...	...
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	AAA	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	AAA	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	AAA	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

# Aperçu de la méthode par étape

- La méthode des EMN vs non-EMN comparables se décline en trois étapes :
  1. **L'identification** de la nature du pays, soit génératrice de l'érosion de la base d'imposition et de transfert de bénéfices, (FFI sortants), soit collectrice de l'érosion de la base d'imposition et de transfert de bénéfices, (FFI entrants) (Voir tableau de bord des indicateurs Erosion de la base d'imposition et le transfert de bénéfices, de l'OCDE).
  2. **La recherche** d'unités pratiquant l'évitement fiscal (TA) parmi les EMN:
    - Les entreprises multinationales italiennes sont évaluées afin de déterminer si elles sont suspectées de comportement d'évasion fiscale, sur la base d'une comparaison entre les entreprises multinationales et un groupe de contrôle composé de non-entreprises multinationales (comparables).
  3. **La vérification** des bénéfices pour les EMN pratiquant l'évitement fiscal:
    - Le ratio BAII/chiffre d'affaires des unités TA est ajusté en explorant le modèle de sélection afin de comparer les résultats économiques des EMN TA par rapport à ceux des EMN non TA.

# Selection – Vue d'ensemble

- La phase de sélection se divise en trois étapes :

- **Définition du groupe témoin**

Pour chaque unité EMN, un groupe témoin d'entreprises nationales est défini en utilisant l'appariement par score de propension.

- **Comparaison entre les entreprises multinationales et les entreprises non multinationales.**

Pour chaque paire unité EMN-groupe de contrôle, une comparaison en termes de part des bénéfices est utilisée pour définir une variable de substitution, qui souligne les éventuels comportements anormaux des EMN.

- **A titre comparatif(parmi les EMN)**

L'analyse ROC est utilisée pour définir le regroupement final entre les unités qui évitent les impôts (TA) et celles qui n'en évitent pas (NTA) à partir de la variable proxy

# Sélection - Définition des groupes témoins

## Définition des variables confusionnelles pour l'analyse d'appariement PS

- **v1** = Rotation / Travailleurs
- **v2** = Nombre d'employés
- **v3** = Coûts des marchandises / Coûts intermédiaires totaux
- **v4** = Valeur des exportations / Chiffre d'affaires
- **v5** = Valeur des importations / Coûts totaux
- **v6** = Salaires / (Salaires + Coûts intermédiaires totaux)
- **v7** = Coût des services / Total des coûts intermédiaires
- **vv** = BAII / Chiffre d'affaires

# Sélection - Définition des groupes témoins

## Analyse de l'appariement par score de propension

- Modèle PS pour définir les probabilités de correspondances

`treat(treated='1') = v1 v2 v3 v4 v5 v6 v7 NUTS2` (Modèle logit pour définir les probabilités de correspondance)

- Méthode d'appariement et nombre de cas similaires

`match method = greedy (k = 5)`

- Caractéristiques de la liaison

`Exact (NACE3 size class NUTS2)`

# Sélection – Comparaison «interne»

## Définition du proxy du suspect

- L'appariement par score de propension permet de définir un groupe de contrôle d'entreprises nationales pour chaque unité EMN.
- Pour chaque paire unité EMN-groupe de contrôle, un **proxy de suspect** de TA est donné par la condition suivante :

- **Proxy = 1**

si le ratio BAII/chiffre d'affaires de l'unité EMN est inférieur à la moyenne du groupe de contrôle.

- **Proxy = 0**

si le rapport entre le BAII et le chiffre d'affaires de l'unité EMN est supérieur ou égal à la moyenne du groupe de contrôle.

# Sélection – Comparaison «interne»

## Étapes de la comparaison interne

- Définition des variables

A partir des caractéristiques structurelles et de performance des entreprises multinationales, on sélectionne les variables qui permettent de cerner le comportement des entreprises multinationales.

- Analyse factorielle

Parmi les variables x0-x8, on distingue deux facteurs.

- Définition de l'indicateur composite.

A partir des facteurs, il définit le composite à utiliser dans le modèle logit de l'analyse ROC.

- Analyse ROC

Il permet de définir la classification finale entre les entreprises multinationales qui évitent l'impôt (TA) et celles qui ne l'évitent pas (nonTA).

# Sélection – Comparaison «interne»

**Définition des variables**((relations inverses avec le proxy, plus les valeurs sont élevées, plus la probabilité de TA est faible).

- **x0** = BAI / Chiffre d'affaires
- **x1** = Valeur ajoutée / Chiffre d'affaires
- **x2** = Coûts de R&D / Total des coûts intermédiaires
- **x3** =  $1 - (\text{Coûts sur les redevances} / \text{Chiffre d'affaires})$
- **x4** =  $1 - \text{Valeur des importations} / \text{Coûts intermédiaires totaux}$
- **x5** =  $1 - \text{Cadre fiscal (par pays, différentiel)}$
- **x6** = Salaires / Chiffre d'affaires
- **x7** =  $1 - \text{Coût des services} / \text{Chiffre d'affaires}$
- **x8** = Valeur des exportations / Chiffre d'affaires

# Sélection – Comparaison «interne»

## L'analyse factorielle et la définition de l'indicateur composite

- Les variables **x0** - **x8** sont d'abord normalisées afin d'éviter les effets d'échelle.
- A partir de l'analyse factorielle, les deux premiers facteurs auto-rotatifs sont retenus et le composite est défini comme suit:

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

où,  $\gamma_{(j,1)}$  et  $\gamma_{(j,2)}$  sont les chargements de la  $j$ -ème variable dans les facteurs 1 et 2,  $x_{(j,i)}$  est la valeur de la  $j$ -ème variable pour la  $i$ -ème observation, et  $\omega_1$  et  $\omega_2$  sont les poids en terme de variance expliquée

	<b>Facteur1</b>	<b>Facteur2</b>
<b>x0</b>	0,043	0,518
<b>x1</b>	0,057	0,519
<b>x2</b>	-0,086	0,025
<b>x3</b>	0,452	-0,063
<b>x4</b>	-0,043	0,178
<b>x5</b>	-0,011	0,050
<b>x6</b>	0,195	-0,002
<b>x7</b>	0,465	-0,049
<b>x8</b>	-0,063	0,118
<b>Variance explicative</b>	2,098	1,688

# Sélection – Comparaison «interne»

## Analyse ROC

- En utilisant le proxy du suspect et l'indicateur composite, le modèle logit suivant peut être exécuté:

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

		Observations	1124	
<b>Profil de réponse</b>				
Proxy	0		Fréquence	237
	1			887
<b>Critère</b>				
			Intersection	Intersection et covariance
	AIC		1159,9	895,9
	SC		1164,9	905,9
	-2 Log L		1157,9	891,9
<b>Résultats</b>				
Paramètres	Estimation		Erreur type	Valeur-P
Composite	-2,4225		0,2075	<.0001
<b>Qualité de l'ajustement</b>				
Pourcentage d'accord	82,4		D di Somers	0,648
Pourcentage de désaccord	17,6		Gamma	0,648

## Sélection – Comparaison «interne»

- L'analyse ROC est liée aux problèmes de classification dans lesquels les classificateurs peuvent donner les quatre résultats possibles indiqués dans **la matrice de confusion**. L'efficacité du classificateur peut être mesurée à l'aide de deux mesures :
- **La sensibilité** mesure la capacité du classificateur à détecter les vrais positifs, i.e.  $TP/(TP + FN)$ ;
- **La spécificité** mesure la capacité du classificateur à détecter les vrais négatifs, c'est-à-dire  $TN/(TN+FP)$ , où elle est généralement considérée dans son expression réciproque (1-Spécificité), qui mesure la détection correcte des faux positifs.

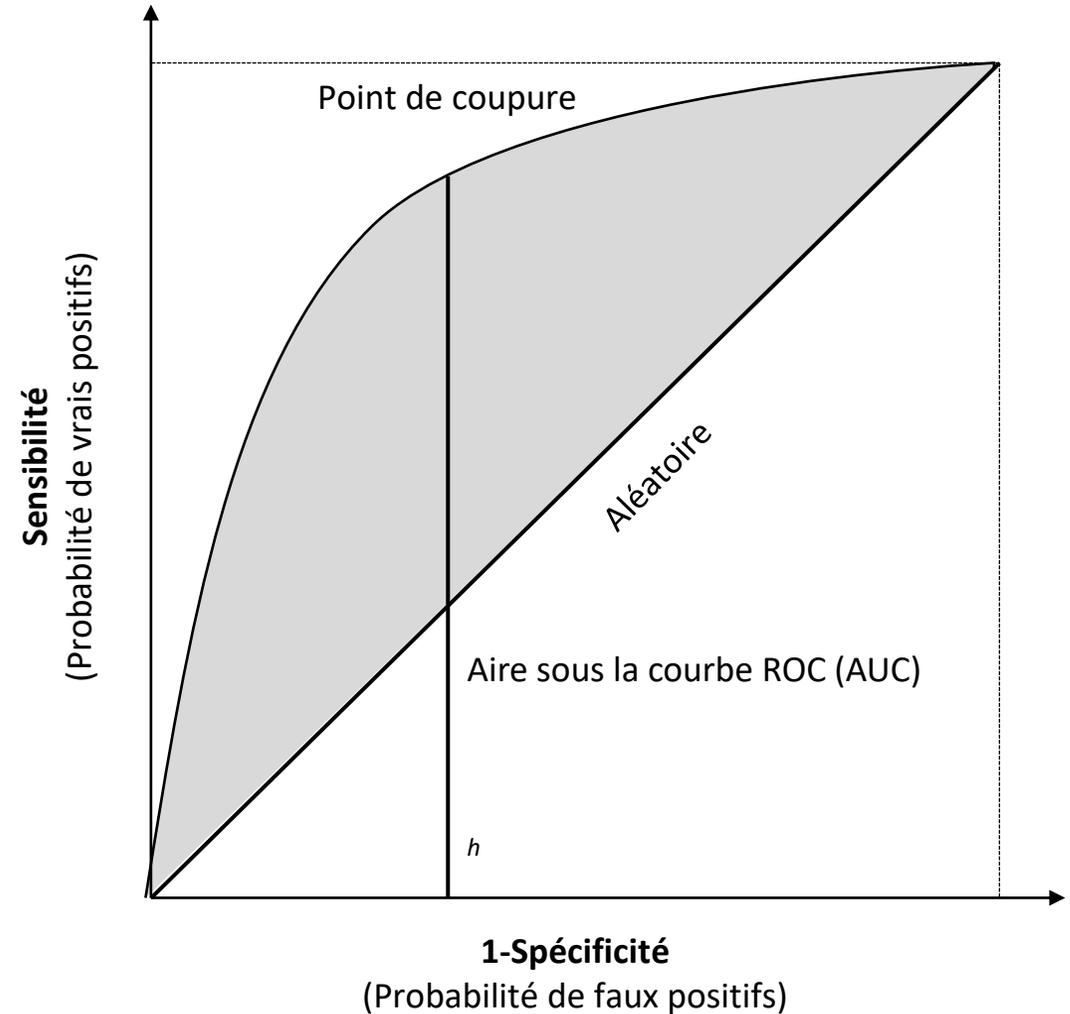
		Classement estimé	
		1	0
Véritable classement	1	<i>TP</i>	<i>FN</i>
	0	<i>FP</i>	<i>TN</i>

# Sélection – Comparaison «interne»

- Considérant un modèle logit ayant :
- un binôme dépendant reflétant un état donné un classificateur représenté par un indicateur unique (même composite)

la distribution des probabilités résultant des estimations logit peut être représentée dans l'espace de la sensibilité et de la spécificité 1 par **la courbe ROC**.

- La ligne de la courbe ROC représente les probabilités attribuées par le modèle à chaque observation dans l'espace du compromis entre la probabilité de détecter des vrais ou des faux positifs pour tous les points de coupure possibles le long des valeurs du classificateur.

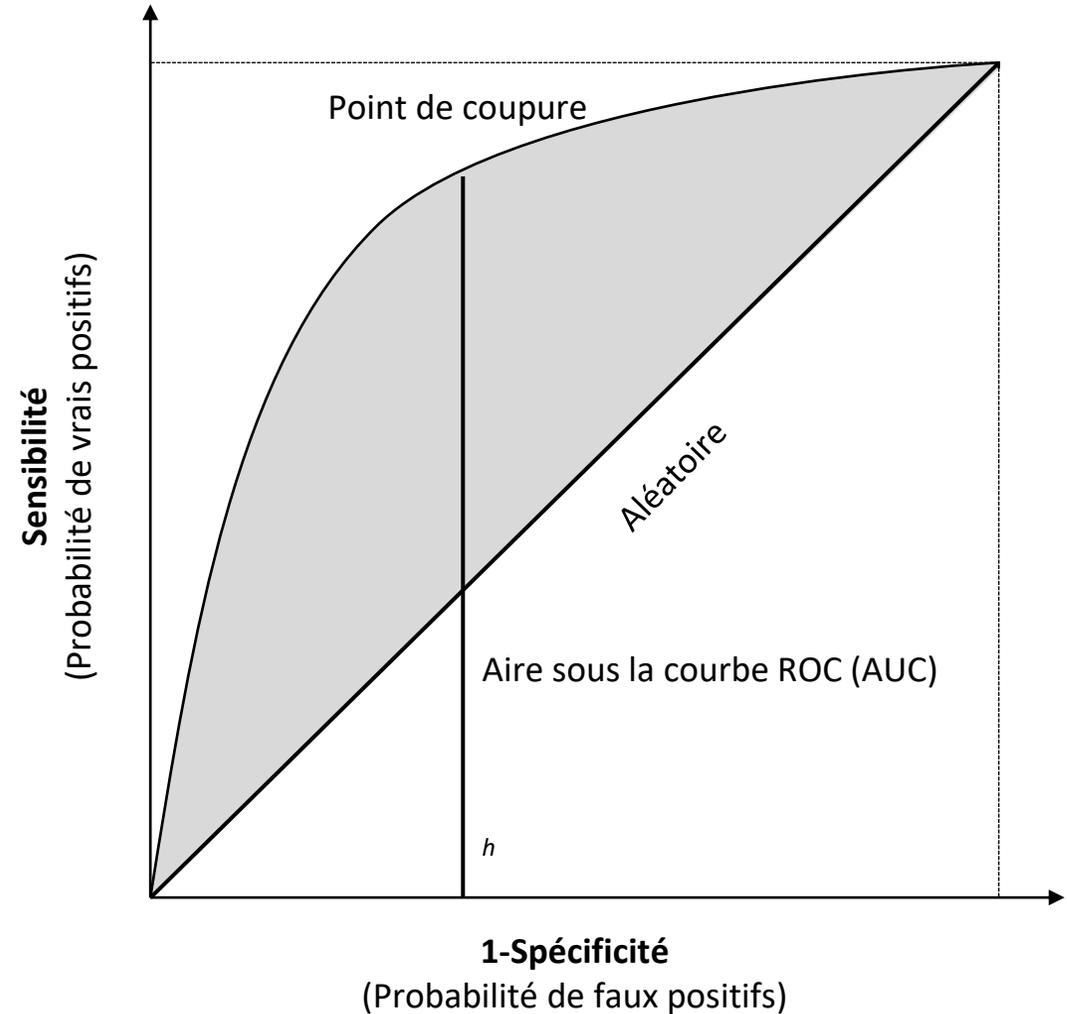


# Sélection – Comparaison «interne»

- Afin d'isoler, le long de la courbe ROC, l'observation qui discrimine le plus efficacement les positifs et les négatifs (Cut), l'équation suivante doit être maximisée:

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

où  $h$  et  $(1-h)$  représentent les poids relatifs permettant de gérer le compromis entre vrais et faux positifs.



# Sélection – Comparaison «interne»

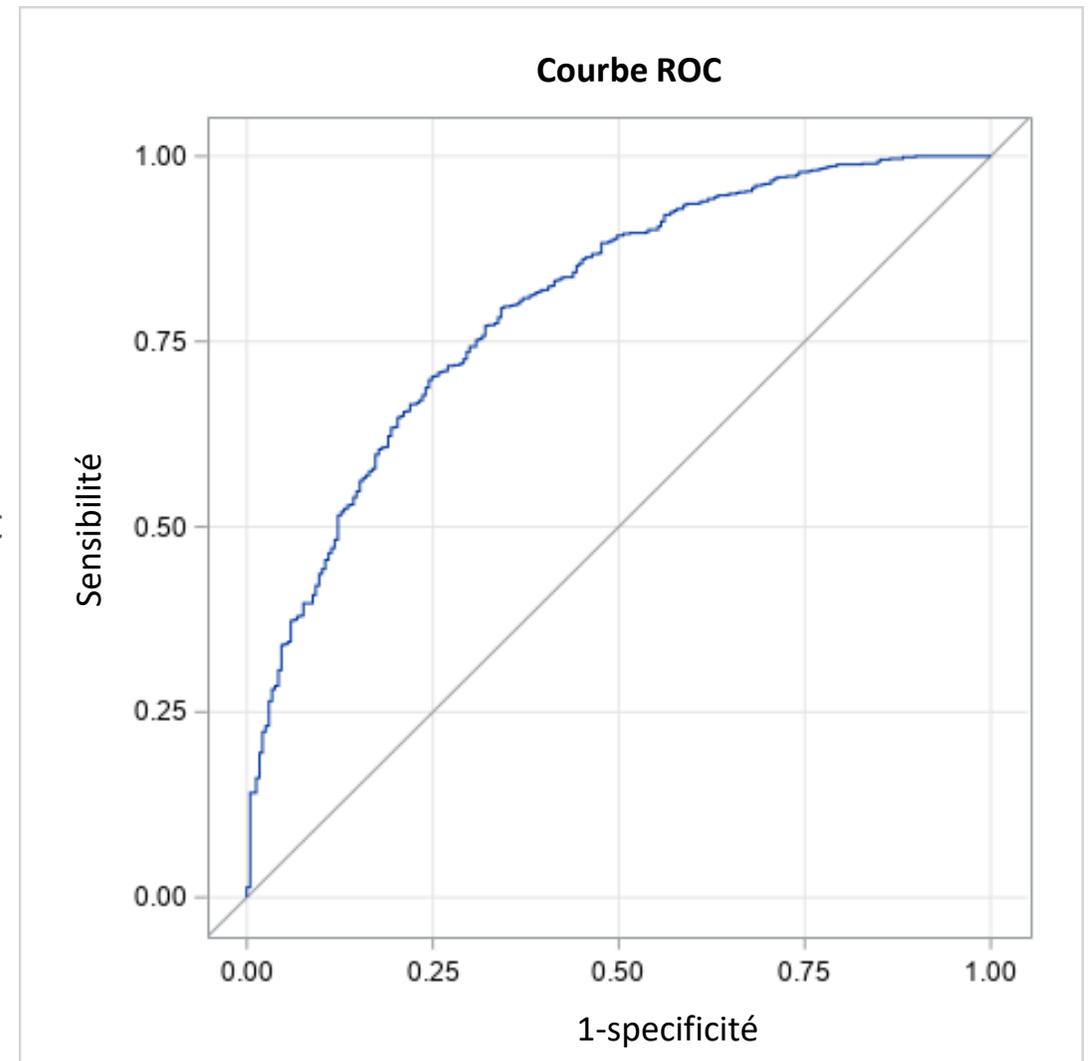
## Analyse ROC

- Le modèle logit génère la courbe ROC suivante (avec AUC=0.8119)

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

avec  $h=0,5$  (sélection neutre), l'ID de l'observation seuil peut être obtenue :

Seuil	ID	Valeur du composite
1	656	0,058488



# Sélection - Comparaison «interne»

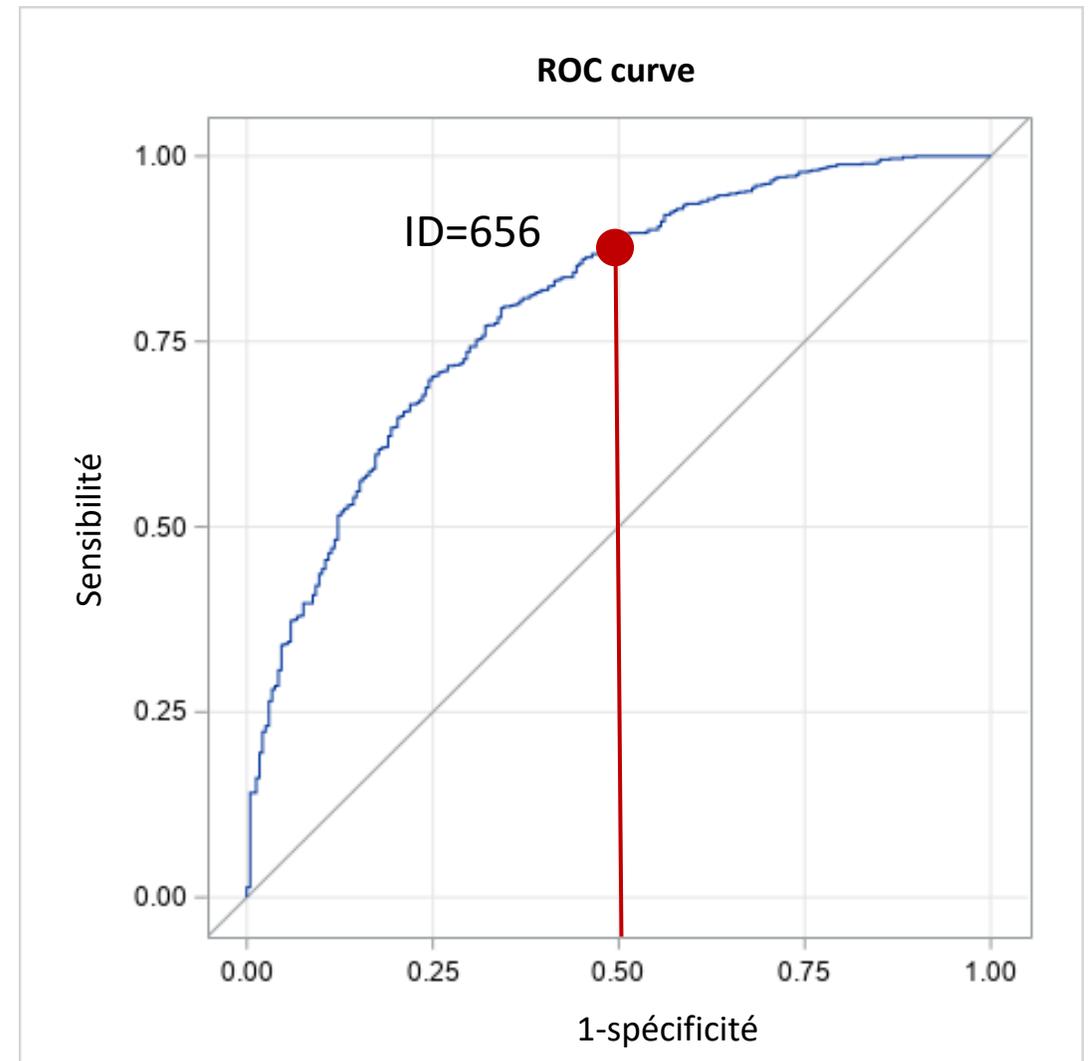
## Classification des EMN

- La valeur du composite du seuil de l'unité MNE  $\bar{S}$  peut être utilisée pour classer d'autres observations.

En particulier:

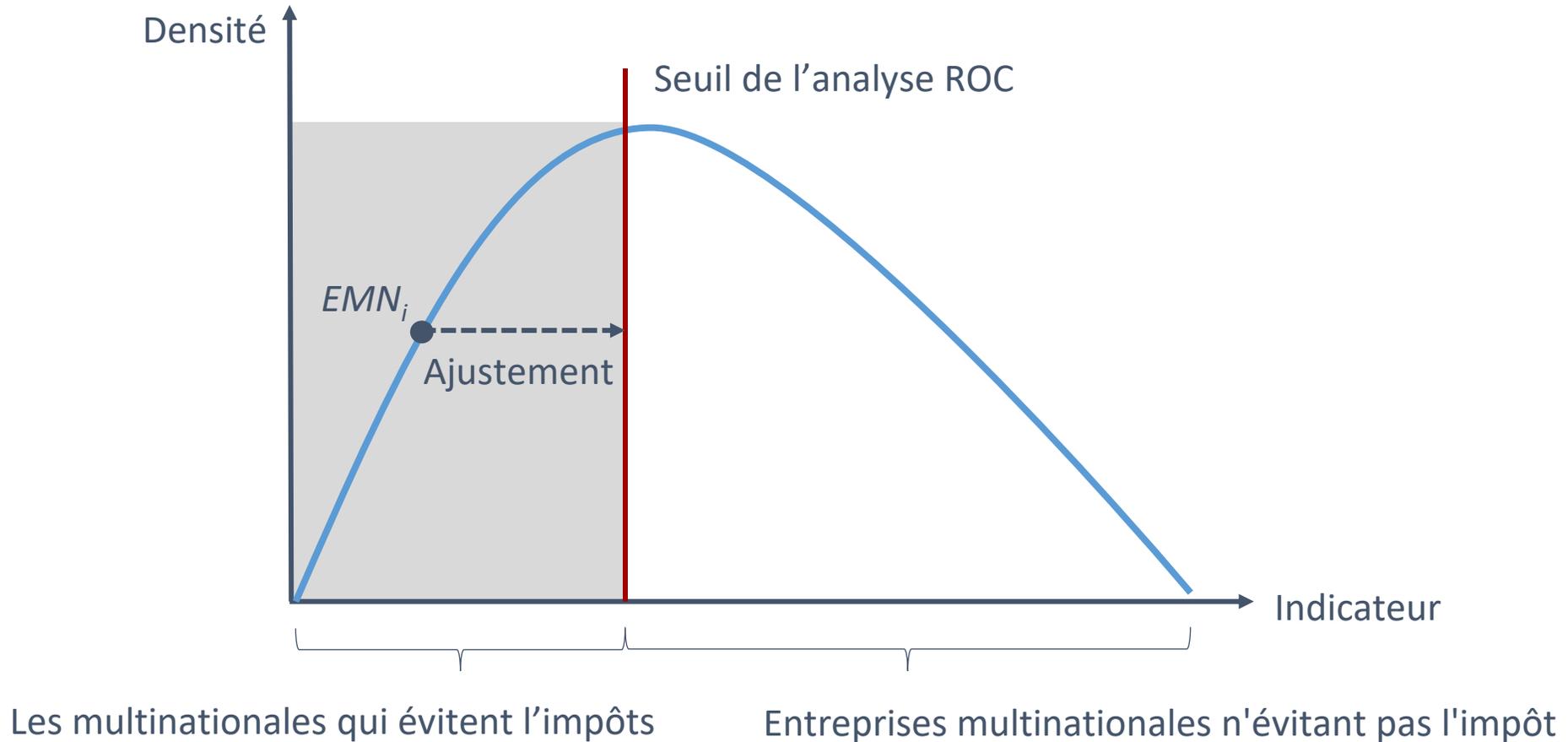
- Si  $I_i \geq \bar{S}$  alors l'unité EMN est non-TA
- Si  $I_i < \bar{S}$  alors l'unité EMN est non-TA

Status		Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
0	Non-TA	441	39,2	441	39,2
1	TA	683	60,8	1124	100,0



# Correction - Cadre conceptuel

- La correction pour l'érosion de la base d'imposition et le transfert de bénéfices, exploite les informations fournies par l'analyse ROC dans la phase de sélection.



# Correction - Cadre conceptuel

- La correction du transfert de bénéfices exploite les informations fournies par l'analyse ROC dans la phase de sélection.
- Pour chaque unité TA, la condition suivante s'applique :

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

où les facteurs sont :

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

- La procédure attribuée à l'indicateur  $x_{_1}$ , qui est le rapport entre le BAII et le chiffre d'affaires, la valeur telle que, pour chaque EMN TA, on obtient la condition suivante :

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

# Correction - Cadre conceptuel

- Cela permet de définir la condition d'ajustement comme :

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

ou:

- $\bar{s}$  est la valeur seuil définie par l'analyse ROC sur l'indicateur composite
- $(\alpha \sum_{-j} \gamma_{-j,1} x_{-j,1} + \beta \sum_{-j} \gamma_{-j,2} x_{-j,2})$  représente l'effet des autres variables sur la valeur de l'indicateur composite
- $\alpha \gamma_{j,1} + \beta \gamma_{j,2}$  représente le poids du rapport entre le BAII et le chiffre d'affaires (CA) sur la valeur de l'indicateur composite.
- $\tilde{x}_{j,i}$  est la valeur ajustée du ratio BAII/CA afin d'amener l'EMN TA sur le seuil.
- Le montant de l'ajustement est obtenu comme suit :  $(\tilde{x}_{j,i} - x_{j,i}) * CA_i$

# Mesure des FFI (sortants et entrants)

- Le montant de la correction est obtenu en comparant le ratio BAII/chiffre d'affaires des deux groupes d'entreprises multinationales définis par le modèle.
- Le montant de la correction représente en fait la mesure des FFI.
- En particulier:

## Pays générateur BEPS

$$FFI_{entrants} = (\tilde{x}_{h,i} - x_{j,i}) * CA_i$$

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

## Pays collecteur BEPS

$$FFI_{sortants}_i = -(\tilde{x}_{h,i} - x_{j,i}) * CA_i$$

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

MERCI.