

Измерение размывания налогооблагаемой базы и вывода прибыли из-под налогообложения (РНБВП): ТНК и сопоставимые не-ТНК

На примере Италии: практика

Федерико Саллюсти

Istat

Построение массива данных

- Для анализа использованы три источника информации:
 - Архив структурной деловой статистики (**Frame-SBS**), который включает структурные и экономические данные по всем 4.4 млн фирм
 - Архив Интегрированной международной базы данных торговли (**COE-TEC**), which includes the information about imports and exports (by product and origin/destination country) for the whole set of firms
 - Архив базы данных **ASIA-Groups** (Итальянская версия Европейского реестра глобальных групп), который включает информацию о компаниях аффилированных с местными и международными группами.
- По каждой единице бизнес-системы итоговая база данных содержит информацию о:
 - Экономической и организационной структуре
 - Характеристиках включения в международную торговлю
 - При наличии, о положении внутри ТНК

Building up the dataset

- Архив структурной деловой статистики (**Frame-SBS**) содержит данные 4.3 млн фирм (2019)
- Архив Интегрированной международной базы данных торговли (**COE-TEC**) содержит данные 4.3 млн фирм в 2019 (165515 интернационализированы, то есть имеют положительные значения импорта/экспорта)
- Архив базы данных **ASIA-Groups** содержит данные 233092 фирм
- Итоговый массив данных нуждается в корректировке (исключение единиц не имеющих сотрудников или добавленной стоимости, отсутствующими данными, представителей отраслей экономики с особенными характеристиками, например табачная промышленность, финансовые учреждения вспомогательного характера и производство и переработка нефтепродуктов)
- Итоговый массив данных содержит 3829908 наблюдений

Построение массива данных

Фирма (ID)	ЕКВЭД	Числ. сотр.	Класс по размеру	Доб.Ст.	Выручка	ФОТ	Промежуточные издержки	Производств. издержки	Затраты на услуги	Лицензионные издержки	НИОКР	Затраты на субподряд	Доходы от субподряда	ЕВИТ	Импорт	Экспорт	Идентификатор группы	Тип группы	Страна головного офиса	Страна нахождения подразделения
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	222	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	222	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	222	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	222	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	222	MUUT	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	222	MUUT	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	222	MUUT	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	222	MUUT	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	222	MUUT	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	222	MUUT	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	222	MUUT	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	222	MUUT	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	222	MUUT	IT	IT

Frame-SBS

COE-TEC

Asia-Groups

Пошаговый обзор метода

- Сравнение ТНК и сопоставимых не-ТНК проводится в три этапа:
 - 1. Установление** факта генерирования (исходящие ТФП) либо притягивания РНБВП (входящие ТФП) в стране(по панели показателей РНБВП стран ОЭСР)
 - 2. Отбор** уклоняющихся от налогов подразделений ТНК
 - По результатам сравнения ТНК с контрольной группой сопоставимых не-ТНК выбираются те ТНК Италии, в отношении которых есть подозрение что они уклоняются от налогов
 - 3. Корректировка** размера прибыли уклоняющихся ТНК
 - Производится корректировка операционной рентабельности подразделений уклоняющейся ТНК чтоб сравнить результаты экономической деятельности уклоняющихся и не-уклоняющихся ТНК

Отбор – Обзор этапа

- **Отбор** производится в три этапа :

Определение контрольной группы

- Для всех подразделений ТНК, определяется контрольная группа местных не-ТНК (используется метод подбора подобного по коэффициенту склонности)

Сравнение между группами (ТНК и не-ТНК)

- В каждой паре ТНК-Контрольная группа проводится сравнение показателей рентабельности и таким образом определяется показатель который будет служить индикатором аномального поведения ТНК

Сравнение в группе ТНК

- Чтоб увидеть возможные кластеры уклоняющихся и не-уклоняющихся подразделений ТНК проводится ROC-анализ на основе косвенной переменной, определенной этапом ранее

Отбор – Определение контрольных групп

Определение переменных для отбора по коэффициенту склонности

- **v1** = Выручка / Численность персонала
- **v2** = Численность персонала
- **v3** = Себестоимость продаж/ Общие промежуточные издержки
- **v4** = Стоимость экспорта/ Выручка
- **v5** = Стоимость импорта/ Общие затраты
- **v6** = Фонд оплаты труда (ФОТ) / (ФОТ + Общие промежуточные издержки)
- **v7** = Стоимость услуг/ Общие промежуточные издержки
- **vv** = Операционная прибыль / Выручка

Отбор – Определение контрольных групп (код)

Отбор подобного по коэффициенту склонности

- Модель подбора

`treat(treated='1') = v1 v2 v3 v4 v5 v6 v7 NUTS2` (Логит-модель для отбора подобного)

- Метод подбора и количество подобных

`match method = greedy (k = 5)`

- Связывающие характеристики

`Exact (NACE3 size class NUTS2)`

Отбор – Сравнение между групп

Определение косвенной переменной, указывающей на подозрение в уклонении от налога

- Методом подбора подобного по коэффициенту склонности определяется контрольная группа местных не-ТНК
- В каждой паре «подразделение ТНК-Предприятие из контрольной группы» определяется **идентификатор** подозрения к уклонению от налога следующим образом:

- **Подозрение = 1**

Если операционная рентабельность подразделения ТНК ниже чем в среднем по контрольной группе

- **Подозрение = 0**

Если операционная рентабельность подразделения ТНК выше или не отличается от средней по контрольной группе

Отбор – Сравнение внутри группы

Этапы сравнения внутри группы

- Определение переменных

Из структурных показателей и показателей эффективности деятельности производится отбор показателей, характеризующих деятельность ТНК:

- Факторный анализ

Из переменных x_0 - x_8 отбирается два фактора

- Определяется комплексный показатель

Из факторов производится комплексный показатель который будет использоваться в логит-модели ROC анализа

- ROC-анализ

Анализ позволяет провести итоговую классификацию ТНК на уклоняющиеся и неуклоняющиеся

Отбор – Сравнение внутри группы

Определение переменных (обратное отношение с косвенной переменной, чем больше значение – тем ниже вероятность уклонения от налогов (УН))

- x_0 = Операционная прибыль / Выручка
- x_1 = Добавленная стоимость (Валовая прибыль) / Выручка
- x_2 = НИОКР / Общие промежуточные издержки
- x_3 = $1 - (\text{Лицензионные издержки} / \text{Выручка})$
- x_4 = $1 - \text{Стоимость импорта} / \text{Общие промежуточные издержки}$
- x_5 = $1 - \text{Налоговая система (по стране, налоговому дифференциалу)}$
- x_6 = ФОТ / Выручка
- x_7 = $1 - \text{Стоимость услуг} / \text{Выручка}$
- x_8 = $\text{Стоимость экспорта} / \text{Выручка}$

Отбор – Сравнение внутри группы

Факторный анализ и определение комплексного показателя

- Сначала переменные **x0 - x8** стандартизируют чтоб убрать влияние эффектов масштаба
- Из факторного анализа оставляются два первых фактора полученных в результате авто-ротации

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

где, $\gamma_{j,1}$ и $\gamma_{j,2}$ это нагрузки переменной j в факторах 1 и 2, $x_{j,i}$ это значение переменной j для i -той единицы, а ω_1 и ω_2 это веса по доле объяснённой дисперсии

	Фактор 1	Фактор 2
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
Объясненная дисперсия	2.098	1.688

Отбор – Сравнение внутри группы

ROC-анализ

Используя косвенный показатель подозрения на уклонение от налога и комплексный показатель можно произвести расчет по следующей логит-модели:

Вероятн. (Подозрение = 1|C)_i = $\Lambda(\alpha C)_i$

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

		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

Отбор – Сравнение внутри группы

- The ROC-анализ может быть представлен классификацией проблем в матрице ошибок **confusion matrix**. Эффективность классификатора определяется по двум параметрам:
- **Чувствительность** оценивает способность классификатора определять истинно-положительные значения, то есть $ИП / (ИП + ЛО)$;
- **Специфичность** оценивает способность классификатора определять истинно-отрицательные значения, то есть $ИО / (ИО + ЛП)$, который обычно представлен показателем обратным своему значению (1 - специфичность), и измеряет насколько верно классификатор определяет отрицательные значения

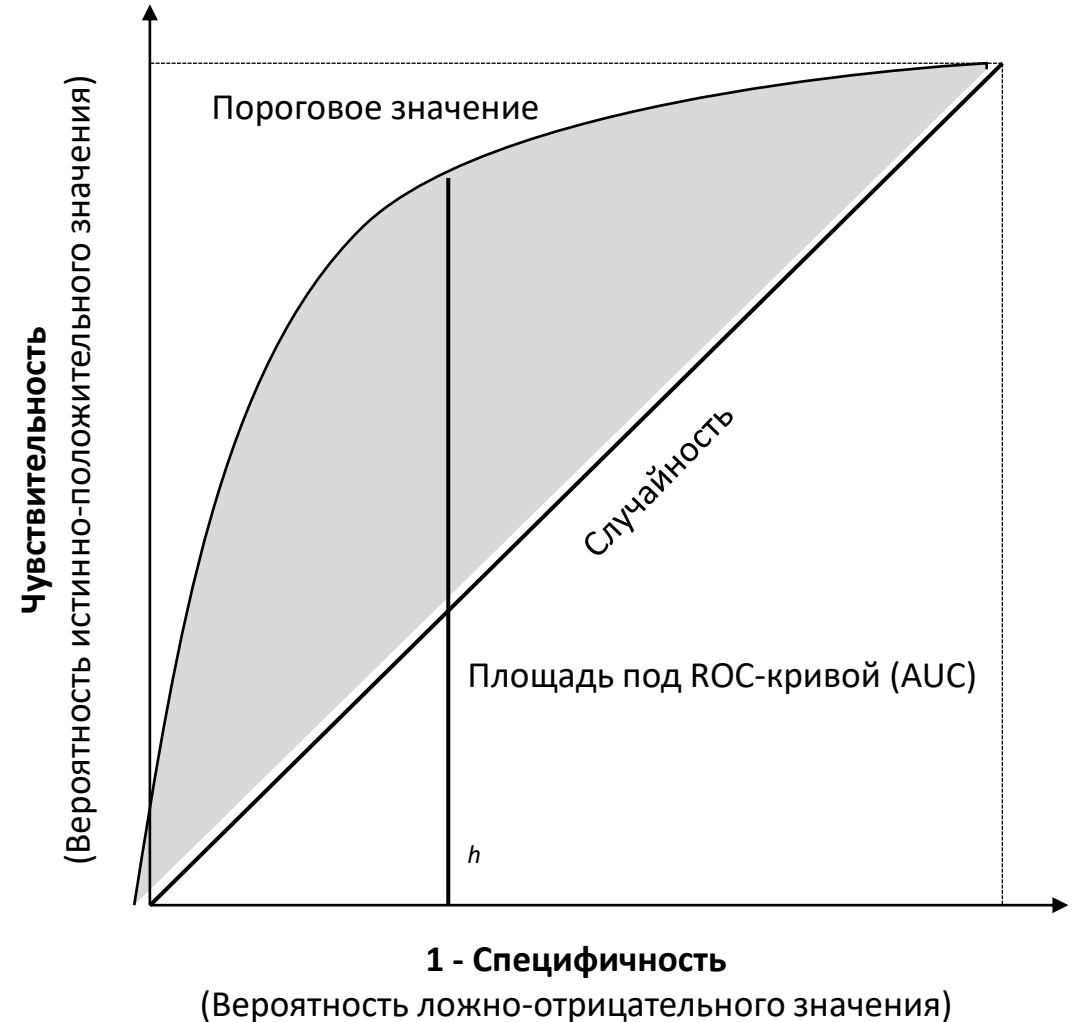
		Оценочная классификация	
		1	0
Истинная классификация	1	<i>ИП</i>	<i>ЛО</i>
	0	<i>ЛП</i>	<i>ИО</i>

Отбор – Сравнение внутри группы

- Учитывая что у логит-модели:
 - Биномная зависимая переменная отражающая состояние
 - Классификатор представлен единичным показателем

, то распределение результатов логит-оценки можно представить в пространстве осей Чувствительности и обратного показателя специфичности по ROC-кривой

- Линия кривой представляет собой вероятностные значения определения истинно-положительных значений по всем пороговым значениям переменной-классификатора, которые модель присваивает каждому наблюдению

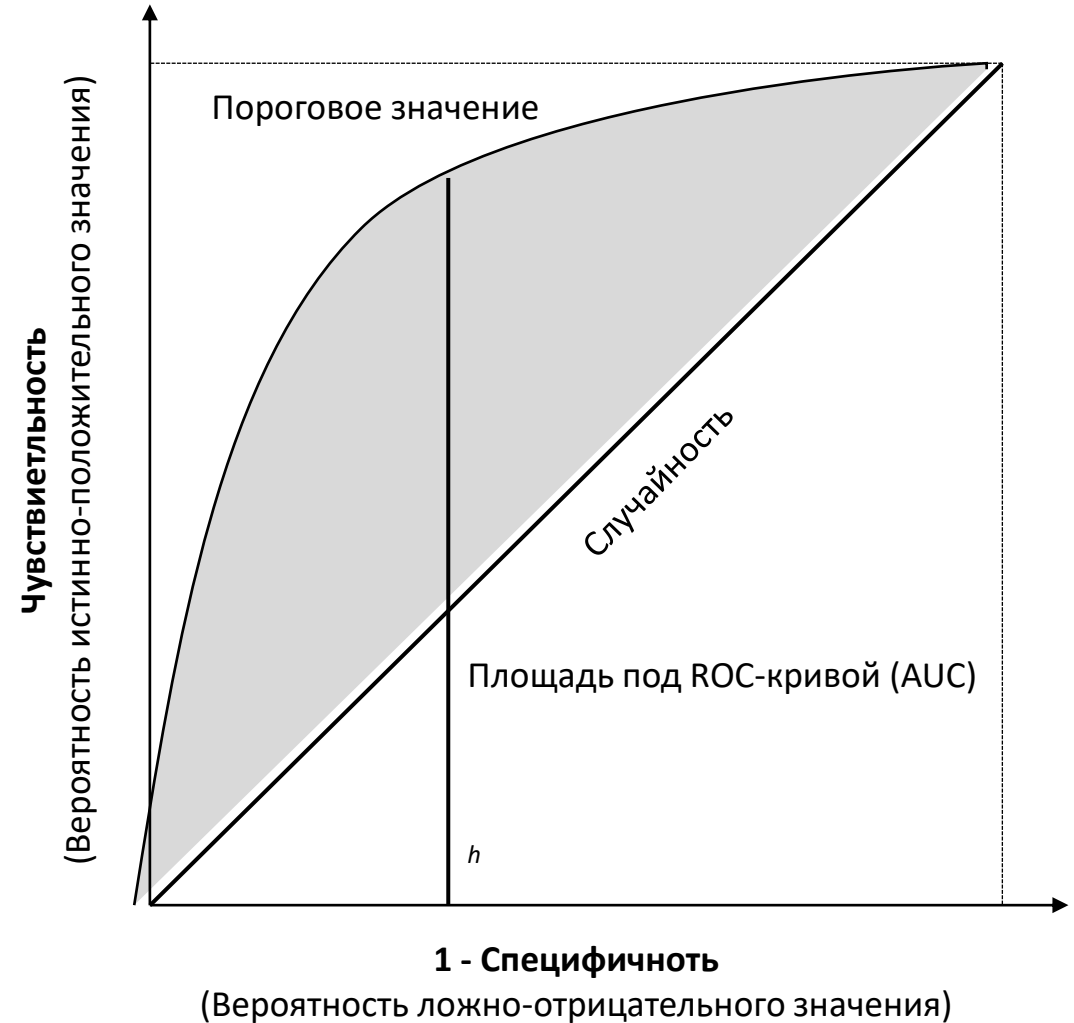


Отбор – Сравнение внутри группы

- Чтобы рассортировать наблюдения по ROC -кривой производится расчет порога, для этого нужно максимизировать следующее уравнение:

$$\text{Порог} = h * \text{Чувствительность} - (1 - h) * (1 - \text{Специфичность})$$

где h и $(1-h)$ представляют собой относительные веса соотношения между истинно и ложно-положительными значениями.



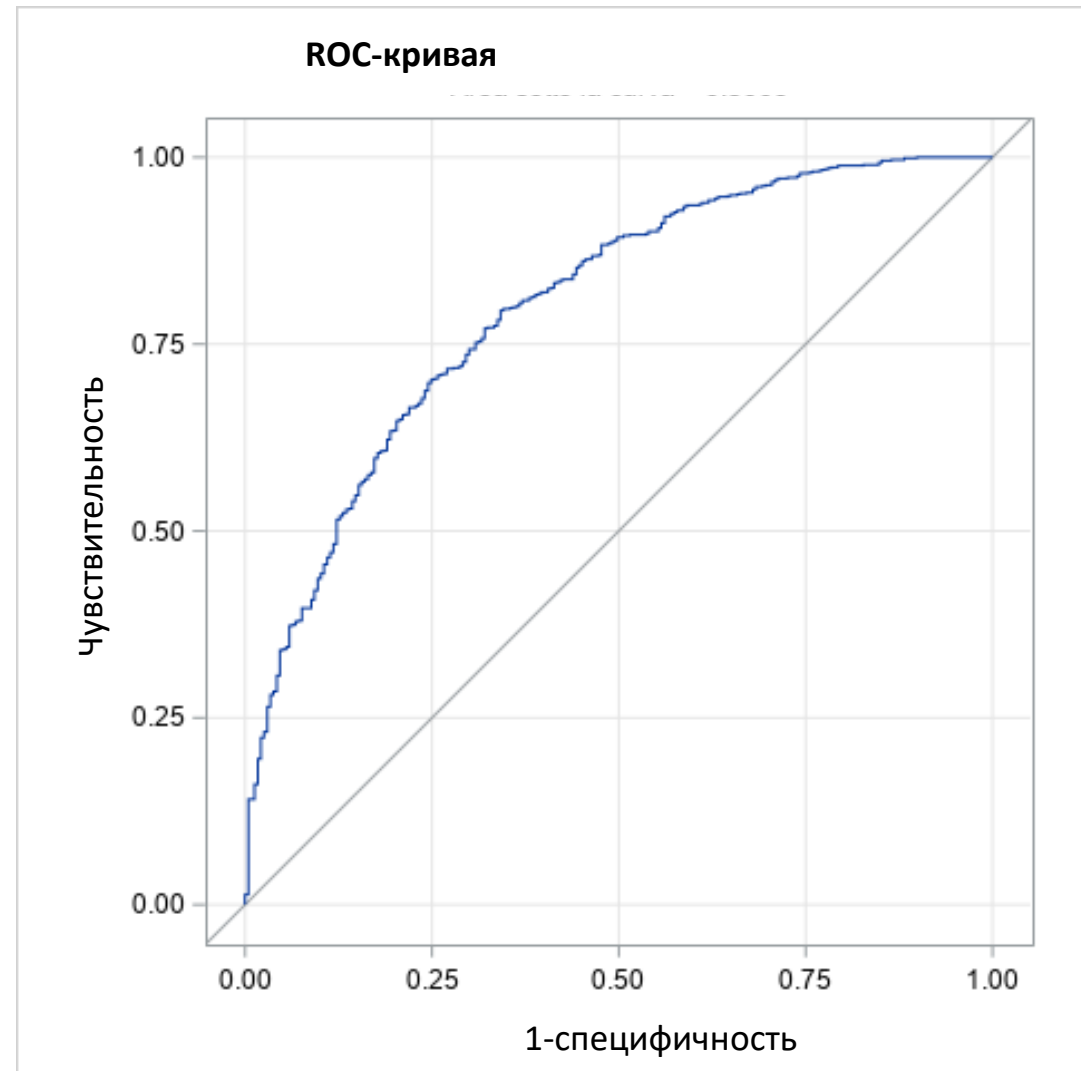
Отбор – Сравнение внутри группы

ROC-анализ

- Логит-модель производит следующую ROC-кривую (с AUC=0.8119 (площадь под кривой))
- Порог = $h * \text{Чувствительность} - (1 - h) * (1 - \text{Специфичность})$

с $h = 0.5$ (нейтральный отбор), получаем следующий идентификатор порогового наблюдения:

Threshold	ID	Value of the composite
1	656	0.058488



Отбор – Сравнение внутри группы

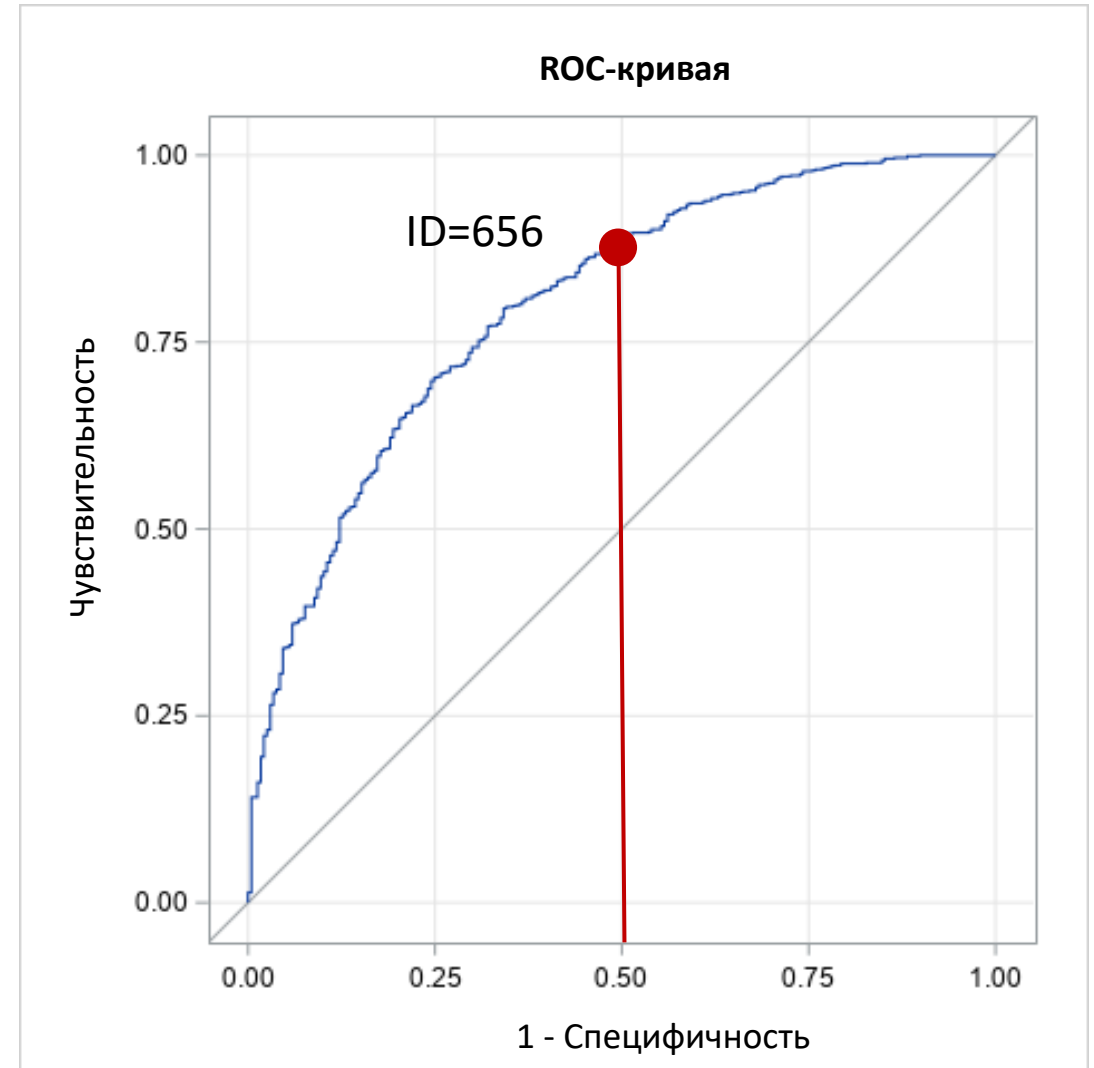
Классификация ТНК

- Пороговое значение комплексного показателя \bar{S} можно использовать для классификации ТНК

В частности:

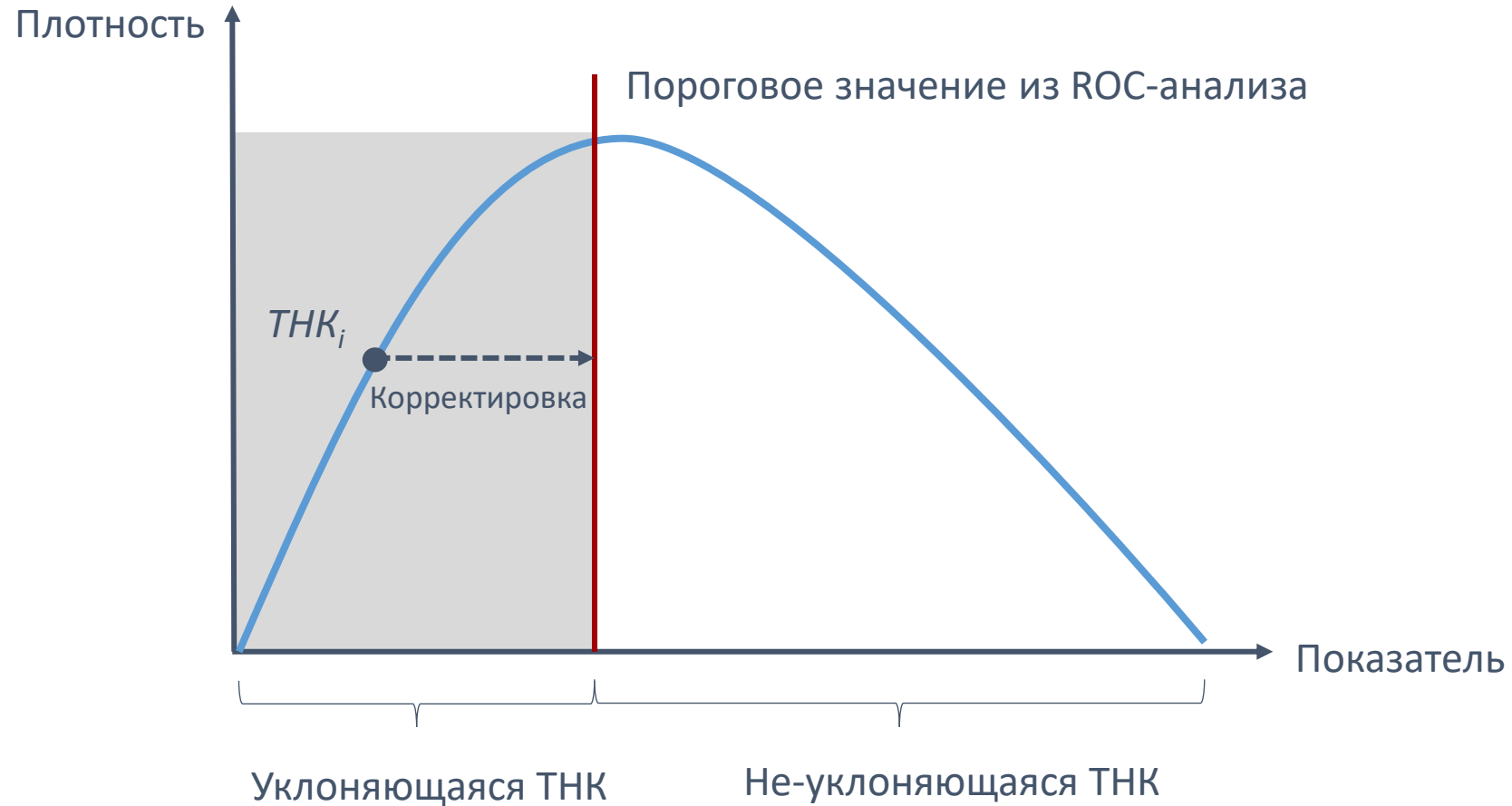
- Если $I_i \geq \bar{S}$ тогда статус ТНК - не-УН
- Если $I_i < \bar{S}$ тогда статус ТНК - УН

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



Коррекция – теоретическая основа

- Корректировка на РНБВП использует информацию этапа отбора



Коррекция – теоретическая основа

- Корректировка на РНБВП использует информацию этапа отбора:
- Для каждого уклоняющегося ТНК верно следующее:

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

, где

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

- Данная процедура присваивает показателю x_1 (операционная рентабельность) каждого уклоняющегося ТНК значение, которое удовлетворяет следующее условие:

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

Коррекция – теоретическая основа

- Таким образом можно определить параметры корректировки как:

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

где:

- \bar{s} это пороговое значение комплексного показателя, определенное ROC-анализом
- $(\alpha \sum_{-j} \gamma_{-j,1} x_{-j,1} + \beta \sum_{-j} \gamma_{-j,2} x_{-j,2})$ это эффект иных переменных на значение комплексного показателя
- $\alpha \gamma_{j,1} + \beta \gamma_{j,2}$ это вес операционной рентабельности в значении комплексного показателя
- $\tilde{x}_{j,i}$ это скорректированное значение операционной рентабельности для того чтоб приблизить уклоняющееся ТНК к пороговому значению
- Величина корректировки определяется как: $(\tilde{x}_{j,i} - x_{j,i}) * \text{Выручка } i$

Измерение (исходящих и входящих) ТФП

- Величина корректировки рассчитывается исходя из сравнения операционной рентабельности двух групп ТНК определённых моделью
- Величина корректировки и представляет собой значение ТФП
- В частности:

Для страны источника РНБВП

$$\text{Исходящие ТФП}_i = (\tilde{x}_{h,i} - x_{j,i}) * \text{Выручка}_i$$

где $\tilde{x}_{h,i} > x_{j,i}$

Для страны объекта РНБВП

$$\text{Входящие ТФП}_i = -(\tilde{x}_{h,i} - x_{j,i}) * \text{Выручка}_i$$

где $\tilde{x}_{h,i} < x_{j,i}$

Спасибо