

# DETECTING UNDER-REPORTING OF VALUE ADDED AND VAT FRAUD IN NATIONAL ACCOUNTS

**Federico Sallusti**

Directorate for National Accounts  
Istat

# Introduction

- Non-observed economy has five components:
  - Underground production
  - Illegal production
  - Informal sector production
  - Household production for own final use
  - Production missed due to deficiencies in data collection program
- Underground production has two components, which can be conceptually separated:
  - Under-reporting, which is the value added not declared by firms in order to reduce tax payments
  - Value added from un-registered workers, which is the value added connected with the use of un-registered workers
- Focusing on under-reporting, the ROC-indicators (ROC-Is) method has been conceived with a twofold purpose
  - To overcome the limitations of the old procedure (“Franz” method)
  - To allow for providing an explicit (instead of implicit) estimate of VAT fraud at micro level

# The old world

- The “Franz” method is based on the comparison between the declared profitability of firms (entrepreneurs) with an opportunity costs proxied by the wage the entrepreneur would be able to gain working as an employee in an analogous production context
- The hidden value added is estimated by imposing that profitability cannot be lower than this opportunity cost for each firm
- This method suffers from three relevant shortcomings:
  - It becomes conceptually unsuitable as the size of enterprises increases (the behavioral assumption implied in the “Franz” method is only suitable if firms are very small in size, i.e. self-employed)
  - It is “by construction” anti-cyclical as wages tend to be less sensitive than profitability with respect to the business cycle
  - The information about structural and economic features of firms is taken into consideration only to a limited extent
- Furthermore, grounding on the comparison between entrepreneurial income and opportunity costs, possible earnings from VAT fraud (without complicity) should be conceptually included in the adjustment, thus losing the possibility to have a separate estimate of under-reporting and VAT fraud

# Towards the new world

- The ROC-Is procedure is an indirect method to detect and adjust under-reporting of value added by firms, whose conceptual framework grounds on the idea that fiscal mis-behaviors can be detected by analysing the incoherence in economic and structural data of firms
- The procedure is based on the application of the ROC analysis, which is used to define the threshold that permits:
  - To classify firms as under-reporting (or not) according to the value of a composite indicator summarising the main economic and structural characteristics of homogeneous firms (economic activity, size and geographical location)
  - To adjust the value added of under-reporting firms based on the same threshold
- This method addresses all the underlined shortcomings of the old-fashioned “Franz” procedure:
  - It maintains its suitability independently from the size of firms as no specific “individual behaviour” is supposed for entrepreneurs.
  - It compares the relative performance of firms, implicitly eliminating any “forced” anti-cyclical behaviour
  - It uses a large set of structural and economic information in the building up of the composite indicator.
- Finally in the ROC-Is method VAT fraud can be separately measured using under-reported value added (calculated at basic price) as tax base

# The ROC-Is procedure (data)

- Data are gathered from the Frame-SBS, an archive integrating survey and administrative sources, which includes economic and structural information for the whole population of Italian productive units (about 4.4 million firms)
- The ROC-Is procedure is applied to the subset of “economically relevant” firms having less than 100 workers and not falling into the conditions of non-treatability (about 2.7 million)
- Indicators relates to three main areas of the economic behavior of productive units:
  - Performance and profitability  
Value added-per-worker; EBITDA-per-self-employed; EBITDA-turnover ratio; Profit-turnover ratio; ROI; Profit-EBITDA ratio
  - Structure of costs  
Labour cost-total cost ratio; Management cost-total cost ratio; Goods/services cost-total cost ratio; Structural cost-total cost ratio; Goods/services costs on inventory rotation
  - Employment structure  
Temporary workers-total workers ratio; Outworkers-total workers ratio; Self-employed workers-total workers ratio)
- In order to keep the homogeneity in terms of the economic and structural characteristics of the firms that are submitted to the ROC-Is procedure, different strata (s) are generated according to the following variables: (1) Industry; (2) Size class; (3) Territory

# The ROC-Is procedure (composite indicator)

- The composite indicator, which should summarise the economic behavior of firm is built in four steps:
  - The binary variable representing the proxy of under-reporting (“suspect”) is defined based on the comparison, for each firm  $i$ , between the suitable performance indicator and the average of the given stratum  $s$ .
  - In the second stage, for each  $s$ , a logit model is estimated having the proxy of “suspect” as dependent variable and the whole set of indicators as covariates
  - A factor analysis is run using the five most informative indicators
  - For each firm  $i$  in the given stratum  $s$ , the three most informative factors are grouped as follows to define the composite indicator:

$$Z_i = \sum_j \omega_j F_{i,j}$$

where  $\omega_j$  are shares of explained variance for each factor, and the  $j$ -th factor is:

$$F_{i,j} = \sum_k \gamma_{k,j} \alpha_{i,k,j}$$

where  $\alpha_{i,k,j}$  is the  $k$ -th indicator for firm  $i$  in the  $j$ -th factor, and  $\gamma_{k,j}$  is the relative loading

# The ROC-Is procedure (identification)

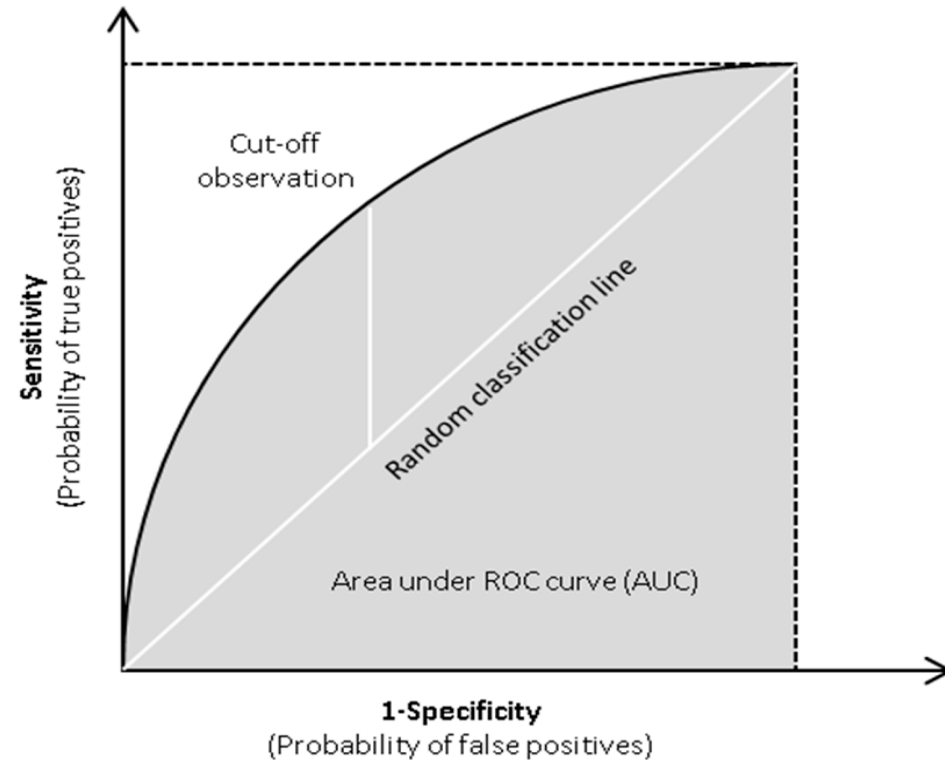
- The identification of under-reporting firms is based on the ROC analysis
- Starting from a logit model, ROC analysis permits to identify a threshold value over the distribution of the covariate, which allows to classify observations with respect to the binary response variable, taking into account the relative weight of possible mis-classification (false positives or negatives)
- In the context of this work, ROC analysis is used to discriminate firms as under-reporting or not based on the relationship between the composite indicator representing the economic behavior of productive units and the proxy representing the “suspect” of fiscal mis-behavior
- With respect to a clustering of firms obtained considering only the results of the proxy vs. composite indicator logit model, the application of the ROC analysis implies two main advantages.
  - It provides a statistical rather than subjective definition of the threshold that clusters positive and negative observations
  - It implicitly corrects the informative capability of the proxy variable by introducing the interaction between the ROC analysis and the economic behavior represented by the composite indicator
- The identification procedure is composed of three steps

# The ROC-Is procedure (identification)

- In the step, the following logit model is estimated:

$$Prob(proxy = 1) = \Lambda(\beta Z_i)$$

- The ROC curve indicates the position of each observation in the space of (the inverse of) specificity and sensitivity, which represent, respectively, the probability of detecting true negatives and true positives using the results of the given logit model.
- Furthermore, the area under ROC curve (the grey area) measures the extent to which the covariate of the logit model is able to discriminate observations with respect to a random selection (represented by the 45° line).





# The ROC-Is procedure (identification)

- In the second stage, the cut-off point over the ROC curve is defined using the following equation:

$$\widehat{Cut} = h * Sensitivity - (1 - h) * (1 - Specificity)$$

where  $h$  and  $(1 - h)$  represent the relative weight assigned to the different elements of the trade-off in clustering

- In particular, when  $h = 0.50$  a “neutral” identification is obtained (the so-called Youden (1950) index). If  $h > 0.5$ , finding true positives is considered more relevant than incurring in false positives. If  $h < 0.5$ , avoiding false positives is considered more relevant than individuating true positives
- Given the value of  $h$ , the permits to identify the cut-off observation along the ROC curve determined by the logit model
- In the third stage, finally, firms are classified as under-reporting or not according to the comparison between the value of their composite indicator and the value of the composite indicator for the threshold observation: assuming  $\bar{Z}$  as the value of the composite indicator for the cut-off observation, other firms are classified as under-reporting if  $Z_i < \bar{Z}$

# The ROC-Is procedure (adjustment)

- The adjustment of value added for under-reporting firms is obtained by exploiting the information provided by the identification stage
- For each under-reporting firm, the adjustment is obtained by increasing the value of the composite indicator (leveraging on the value added-per-worker indicator) up to the threshold value defined by the ROC analysis
- This way, each formerly under-reporting firm is brought to the threshold value in order to be considered as not under-reporting
- The adjustment is thus obtained using the following condition:

$$\tilde{\alpha}_{h,i}: \sum_j \omega_j F_{i,j} = \bar{Z}$$

# The ROC-Is procedure (adjustment)

- Using algebra, the following can be obtained:

$$\tilde{\alpha}_{h,i} = \frac{\bar{Z} - \sum_j \omega_j \gamma_{j,-h} \alpha_{-h,i}}{\sum_j \omega_j \gamma_{j,h}}$$

where, the magnitude of the adjustment will depend on:

- The value of  $\bar{Z}$ , which represents contextual conditions at industry level
- The magnitude of the effect of the indicators other than the value added-per-worker  $\sum_j \omega_j \gamma_{j,-h} \alpha_{-h,i}$
- The magnitude of the effect of the value added-per-worker indicator  $\sum_j \omega_j \gamma_{j,h}$
- Overall, the amount of adjustment  $y_i = (\tilde{\alpha}_{h,i} - \alpha_{h,i}) * N_i$  will be directly connected to the distance between the value of the composite indicator for under-reporting firms and the threshold, which indicates to what extent the given firm is “abnormal” with respect to the rest of the stratum
- Note that  $N_i$  is the number of registered workers (this is why estimates of under-reporting and estimates of value added coming from the use of un-registered workers can be independently carried out)

# Measuring VAT fraud

- Following the EU Commission decision on the treatment for national accounts purposes of the non-collected VAT, VAT fraud has to be included in GDP and GNI according to the characteristics of the transactions generating it:
  - VAT fraud with complicity, which is generated in transactions in which parties consensually decide to hide the tax base
  - VAT fraud without complicity, which is generated in transactions in which the seller hides the tax base, while the purchaser pays the related VAT
- Only VAT fraud without complicity should be included in National Accounts
- In order to provide an estimate of VAT fraud, two main conditions have to be assured
  - The suitable tax base should be measured
  - The two typologies of fraud (with and without complicity) should be distinguished
- Considering the first issue, while in “Franz” method, the adjustment was connected with the total flow of resources entering in the firm, thus also including possible VAT fraud without complicity, in ROC-Is method only value added at basic price is taken into account, thus explicitly excluding the possibility to overlap under-reporting and VAT fraud
- Considering the second issue, using administrative data on VAT payments provided by the Italian tax authority, and making few assumptions, it is possible to distinguish between the amount of adjusted value added connected with the two typologies of VAT fraud

# Measuring VAT fraud

- The procedure to estimate VAT fraud without complicity relies on two main assumptions:
- The first assumption relates to the fact that fraud without complicity is more likely to occur in business-to-consumers (B2C) transactions, while parties have higher incentives to carry out VAT fraud with complicity in business-to-business (B2B) transactions.
- This main assumption is relaxed in order to take into account two possible issues:
  - Cases in which also in B2B transactions VAT fraud without complicity can occur are coped with taking into account the (sectoral) share of mismatch between outward and inward invoices coming from the buyer-supplier database of the Italian tax authority
  - Cases in which complicity can occur also in B2C transactions are dealt with by identifying specific markets (industries) in which consumers and firms can have the possibility and the incentive to fraud tax authority consensually (e.g. professionals, constructions)
- The second assumption claims that the distribution between B2B and B2C transactions for the under-reported value added is the same as the distribution for the declared value added

# Measuring VAT fraud

- The procedure to estimate VAT fraud is composed of five steps
- In the first stage of the procedure, using administrative microdata about VAT payment of firms provided by the Italian tax authority, for each productive unit  $i$ , the following are defined:

$$y_{B2B,i} = \mu_{B2B,i} * Y_i$$
$$y_{B2C,i} = \mu_{B2C,i} * Y_i$$

where  $y_i$  is the adjustment for firm  $i$  and  $\mu$ 's are the share of declared value added coming from B2B and B2C transactions

- In the second stage, using the same dataset, implicit VAT rates for B2B and B2C are calculated:

$$\rho_{B2B,i} = T_{B2B,i}/Y_{B2B,i}$$
$$\rho_{B2C,i} = T_{B2C,i}/Y_{B2C,i}$$

where  $T$ 's are the amount of VAT paid by the firm  $i$  for the relative typology of transaction, and  $Y$ 's are the amounts of declared value added by typology of transaction

# Measuring VAT fraud

- In the third stage, the amount of VAT fraud related to under-reporting of value added is then determined by multiplying the tax bases the implicit VAT rate:

$$t_{B2B,i} = \rho_{B2B,i} * y_{B2B,i}$$

$$t_{B2C,i} = \rho_{B2C,i} * y_{B2C,i}$$

- In the fourth stage  $t_{B2B,i}$  and  $t_{B2C,i}$  are corrected in order to take into account, at industry level, the possibility of VAT fraud with complicity in B2C and Vat fraud without complicity in B2B
  - The first case is dealt with by assuming, for each industry, a share of B2C transactions that generates VAT fraud with complicity

$$\bar{t}_{B2C,s} = \rho_{B2C,s} * y_{B2C,s} * \tau_{B2C,s}$$

- The second case is dealt with by considering, for each industry  $s$ , the share of this type of transaction that are suspected of VAT fraud without complicity

$$\bar{t}_{B2B,s} = \rho_{B2B,s} * y_{B2B,s} * \tau_{B2B,s}$$

# Measuring VAT fraud

- In the fifth stage, the total amount of VAT fraud by typology is obtained as follows:

$$t_C = \sum_s \left( \sum_i t_{B2B,i} \right) - \bar{t}_{B2B,s} + \bar{t}_{B2C,s}$$

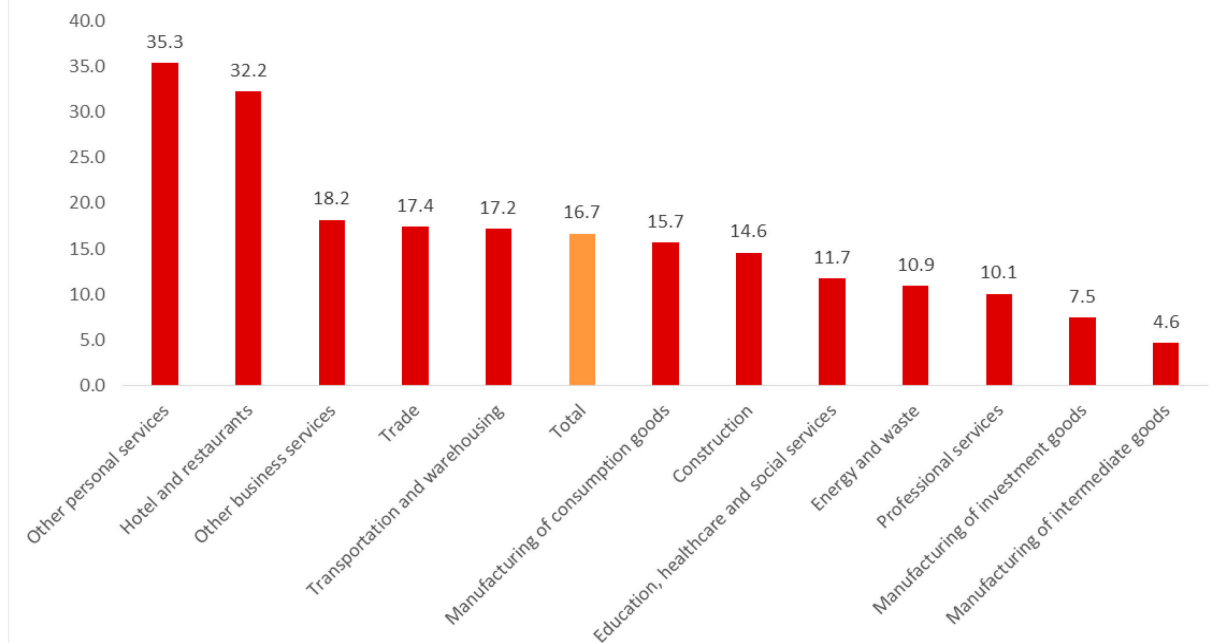
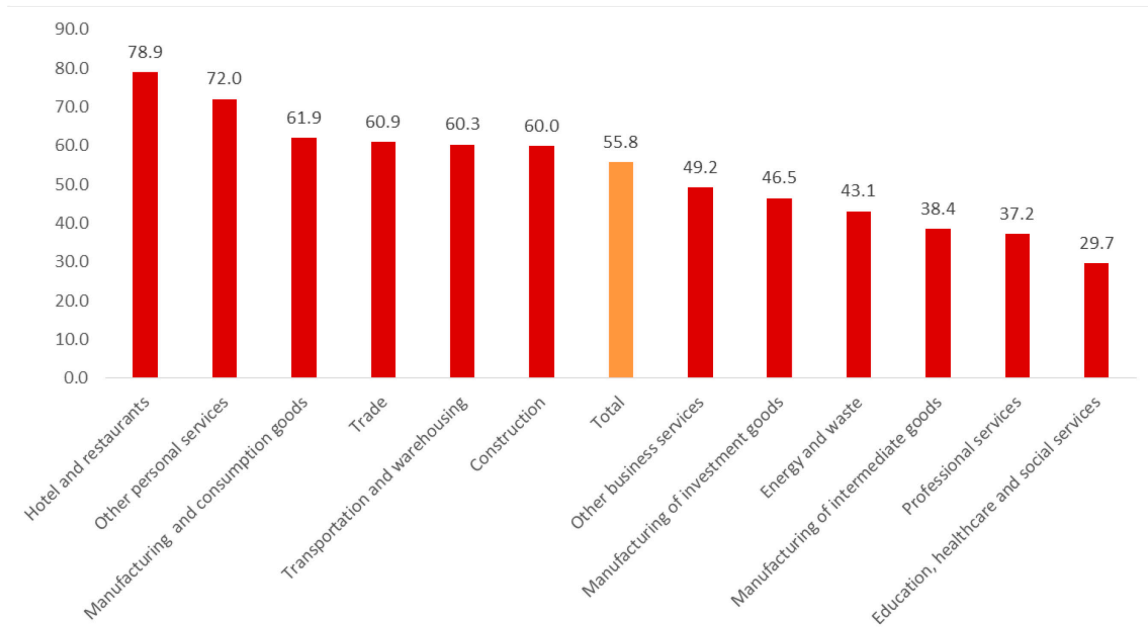
$$t_N = \sum_s \left( \sum_i t_{B2C,i} \right) - \bar{t}_{B2C,s} + \bar{t}_{B2B,s}$$

where  $t_C$  is the final amount of VAT fraud with complicity and  $t_N$  is the final amount of VAT fraud without complicity to be included in the GDP



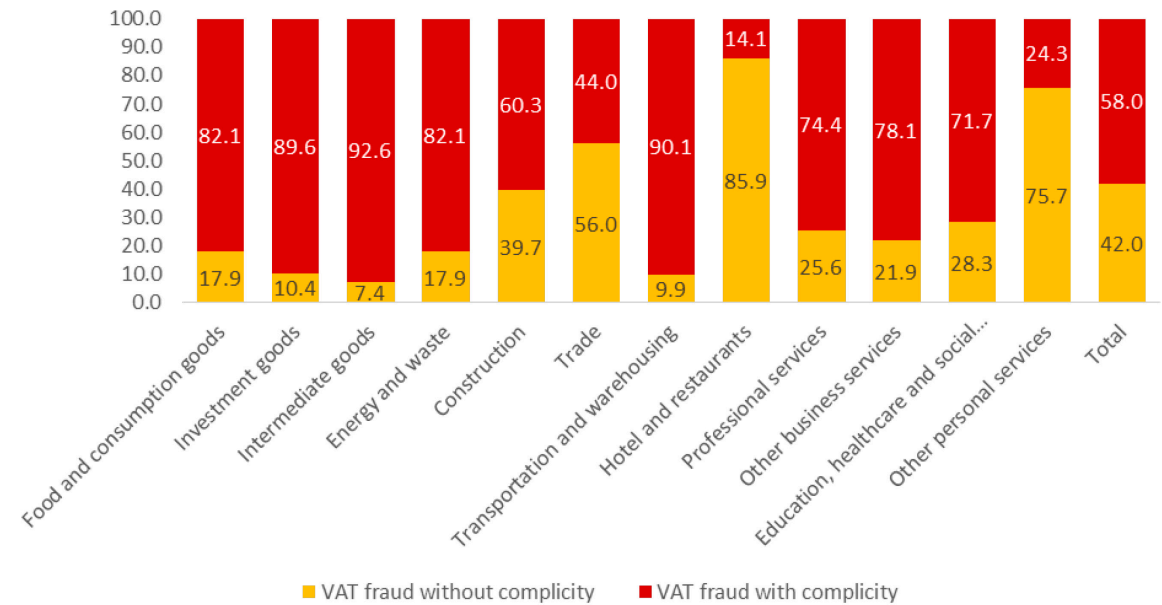
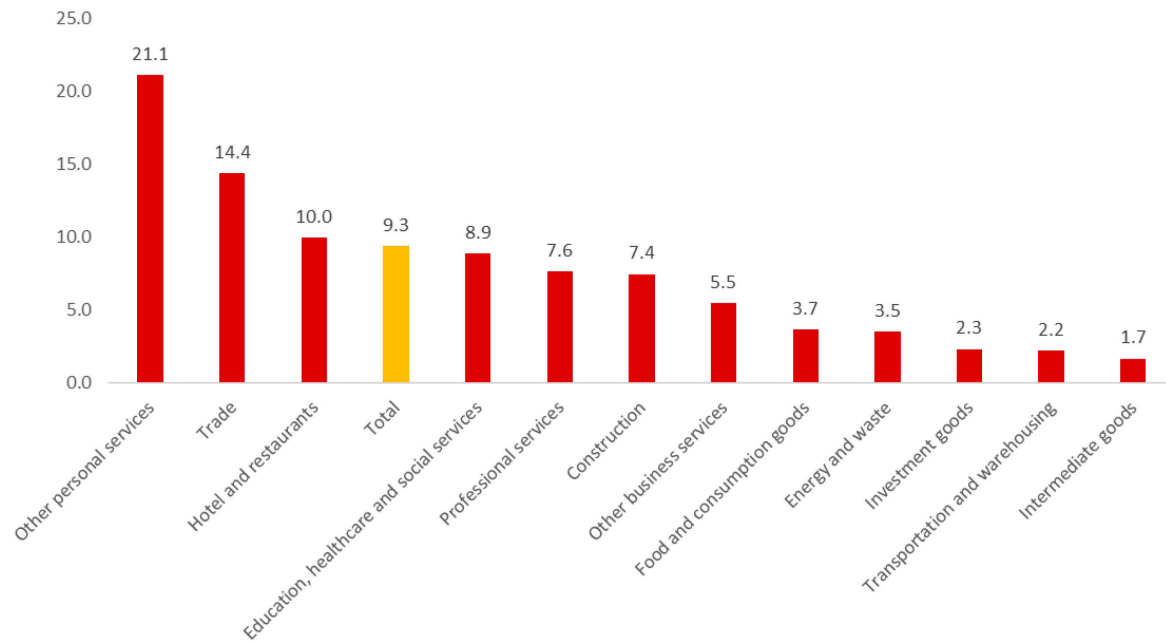
# Results

- The adjustment for under-reporting related to the sub-population of firms for which the ROC-Is method is applied amounts to 64.7 billion euro in 2016, representing about 80% of the total adjustment
- On average, the correction impacts for about 20% with respect to the declared value added, while the final incidence of under-reporting with respect to the adjusted value added is over 16%
- Under-reporting firms are about 55% of the whole sub-population



# Results

- VAT fraud without complicity related to under-reporting is about 4.5 billion euros, and represents about the 80% of the total
- On average VAT fraud without complicity accounts for 9.3 percent of total adjustment
- On average VAT fraud without complicity represents 42.0 percent of total VAT fraud related to under-reporting



# Conclusion

- The new ROC-Is method has been conceived with the goal of overcoming the set of limitations of the old “Franz” procedure:
  - Conceptual reliability is independent from the size of firms
  - Results are not forcedly anti-cyclical
  - A large stream of the micro-level information contained in the Frame-SBS archive is exploited
  - VAT fraud without complicity, which has to be included in GDP and GNI, can be separately estimated
- In ROC-Is procedure, both identification and adjustment are conceptually linked to the capability of classifying firms based on the positioning of each firms with respect to a threshold
- Firms are classified taking into account both their peculiar characteristics and the general trend and structure of the industry, avoiding any subjective evaluation.
- The possibility to measure VAT fraud separately is a substantial improvement with respect to the old procedure, in which VAT fraud was included in the total adjustment without any further information about its magnitude.

Thank you.  
fsallusti@istat.it