United Nations ECE/CES/GE.20/2022/9

Economic and Social Council

Distr.: General
13 April 2022

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on National Accounts

Twenty-first session
Online, 17, 19, 23 and 25 May 2022
Item 5 of the provisional agenda
Informal economy

Detecting under-reporting of value added and value-added tax fraud in National Accounts

Prepared by Italian National Institute of Statistics*, **

Summary

This document introduces the recently developed procedure by the National Institute of Statistics of Italy (ISTAT) for improving the measurement of under-reporting of value added and value-added tax (VAT) fraud. The new measurement procedure has been developed with a twofold purpose: to overcome the limitations of the old procedure for measuring under-reporting and to allow for providing a point estimate of VAT fraud at micro level. It is based on the application of the new Receiver Operating Characteristics (ROC) Indicator method, which uses the wide informative set provided by the "Frame-Structural Business Statistics" containing comprehensive economic and structural information for the whole population of Italian firms (about 4.4 million units). The ROC analysis permits to classify firms as under-reporting or not, according to the value of a composite indicator, which summarizes the main economic and structural characteristics of homogeneous enterprises. The same threshold is then used to adjust the value added of under-reporting firms. Furthermore, using the ROC indicator method for measuring under-reporting also allows for providing a separate firm-level estimate for VAT fraud.

* Prepared by Federico Sallusti and Luciano Cavalli, Directorate for National Accounts. The opinions expressed in this work and articles are those of the authors and do not involve the responsibility of the National Institute of Statistics.

** Late submission for translation due to delayed receival of the document.

I. Introduction

1. According to the last estimates the National Institute of Statistics of Italy (ISTAT, 2018), the Non-Observed Economy (NOE, hereinafter) accounts for about 12.4% of the gross domestic product (GDP) and 13.8% of the whole amount of value added generated by the Italian business system in 2016.

2. OECD (2002) defines five components of NOE: (1) underground production; (2) illegal production; (3) informal sector production; (4) household production for own final use; (5) production missed due to deficiencies in data collection program.

3. In Italy, underground production, which is mainly connected with the willingness to avoid the payment of tax and social contribution, is by far the most relevant of them, representing more than 90% of the total amount of NOE in 2016.

4. Underground production has two components, which, although interrelated in practice, can be separated conceptually. Namely, on the one hand, hidden value added can be connected with the false declarations of firms aimed at under-reporting their value added (i.e. under-report production or over-report costs) in order to reduce tax payments. On the other hand, hidden value added can be related to the employment of workforce that is not registered in order to avoid the payment of social contribution.

5. This work focuses on the first component. In particular, the paper presents the Receiver Operating Characteristics indicators (ROC-Is) method recently developed by ISTAT for improving the measurement of under-reporting of value added and value-added tax (VAT) fraud at micro level.

6. This new procedure has been conceived with a twofold purpose: (1) to overcome the limitations of the old procedure; (2) to allow for providing an explicit (instead of implicit) estimate of VAT fraud.

7. Before the last benchmarking activities for the introduction in 2014 of the European System of National and Regional Accounts (ESA 2010) standard for the compilation of National Accounts, under-reporting of value added was estimated by using a procedure grounding on the so-called “Franz method” (Franz, 1983).

8. In this context, under-reporting of value added is measured by comparing 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.[1] In particular, hidden value added is estimated by imposing that profitability cannot be lower than this opportunity cost for each firm.

9. This method suffers from three relevant shortcomings. First, it becomes conceptually unsuitable as the size of enterprises increases. Indeed, the behavioural assumption implied in the Franz method is only suitable if firms are very small in size (i.e. self-employed). Second, it is “by construction” anti-cyclical as wages tend to be less sensitive than profitability with respect to the business cycle. Third, the information about structural and economic characteristics of firms is taken into consideration only to a limited extent.

10. 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.

11. The new ROC-Is method uses the wide informative set provided by the so-called Frame-SBS database, which includes comprehensive economic and structural information for the whole population of Italian firms (about 4.4 million units).

[1] In other words, given a set of economic assumptions (i.e. competitive market, flexible labor market, perfect fungibility between dependent and independent jobs), the entrepreneur would not rationally accept to earn from his firm less than what he could earn by working as employee.

12. The procedure is based on the application of the ROC analysis, which is widely used in medicine and machine learning and has recently become a relevant tool in economics (Costa *et al.*, 2019). In particular, ROC analysis is used to define the threshold that permits to classify firms as under-reporting (or not) according to the value of a composite indicator, which summarises the main economic and structural characteristics of homogeneous firms in terms of economic activity, size and geographical location. The same threshold is then used to adjust the value added of under-reporting firms.

13. This method addresses all the underlined shortcomings of the old-fashioned "Franz" procedure. First, it maintains its conceptual suitability independently from the size of firms as no specific "individual behaviour" is supposed for entrepreneurs. Second, it compares the relative performance of firms, implicitly eliminating the "forced" anti-cyclical behaviour of the "Franz" method. Third, it uses a large set of structural and economic information in the building up of the composite indicator.

14. Finally, using the ROC-Is method for measuring under-reporting also allows for providing a separate estimate for VAT fraud without complicity. Indeed, in the new procedure, VAT fraud can be separately measured using under-reported value added (calculated at basic price) as tax base. The amount of VAT fraud is then calculated by applying, at micro level, the share of the VAT paid by regular enterprises (in the same domain) to the amount of value added evaded by under-reporting firms.

15. The remaining of the work is organised as follows. The second section is aimed at presenting the ROC-Is procedure and the database used for the analysis. The third section is finalised to describe the procedure allowing to estimate VAT fraud starting from the results of the measurement of under-reporting. The fourth section shows the results obtained for 2016. The fifth section concludes.

II. Receiver Operating Characteristics -indicators procedure

16. The magnitude and characteristics of non-observed phenomena can be hardly measured using direct approaches. Generally, they are indeed approached applying indirect estimates, which ground on the information coming from observable data, with the aim of identifying "abnormal" behaviours and, possibly, attributing this deviation from "normality" to the non-observed phenomenon.

17. In this context, reliability and suitability of indirect approaches depends on the validity of the conceptual framework used to interpret un-observed phenomena, and on the characteristics of the data that are used to measure them. The former is crucial in defining if, and to what extent, the deviation from normality is attributable to the non-observed behaviour, while the latter is relevant in assuring the coherence between the conceptual model and its information counterpart.

18. The ROC-Is procedure is an indirect method to detect and adjust under-reporting of value added by firms. The conceptual framework grounds on the idea that fiscal misbehaviours can be detected by analysing the incoherence in economic and structural data of firms. In this context, the ROC analysis allows for determining the threshold over which incoherence can be probabilistically attributed to under-reporting.

19. The rest of the section is organised as follows. The first subsection is devoted to stress manipulation and processing of data. The second subsection shows the different stages of the ROC-Is procedure.

A. Data

20. The possibility to define a procedure for measuring under-reporting at micro level implies the availability of a large set of information about structural and economic characteristics of firms.

21. In the last years, ISTAT developed Frame-Structural Business Statistics (SBS) (Monducci and Luzi, 2014), 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).

22. The ROC-Is procedure is applied to the subset of "economically relevant"[2] firms having less than 100 workers and not falling into the conditions of non-treatability[3] (about 2.7 million firms in 2017).

23. In order to describe the economic behavior of firms, a large set of indicators have been defined and gathered from the Frame-SBS archive. In particular, for each firm, indicators have been normalized to have a monotonic relationships with the suspect of under-reporting.

24. Indicators relates to three main areas of the economic behavior of productive units:

- **Performance and profitability**: Value added-per-worker; EBITDA[4]-per-self-employed; EBITDA-turnover ratio; Profit-turnover ratio; Return on investment; 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.

25. 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.

B. The Receiver Operating Characteristics -Is procedure

26. The ROC-Is procedure is composed of three stages. (1) The first is represented by the construction of the composite indicator that summarises the economic behaviour of firms. (2) The second is aimed at defining, using ROC analysis, the threshold which permits to classify firms as under-reporting or not (i.e. identification). (3) The third is finalised to determine the amount of the adjustment that is needed in order to make the value added of under-reporting firm coherent with the condition of absence of fiscal misbehaviour (i.e. adjustment).

1. Economic behaviour: the composite indicator

27. The ROC analysis allows to define a cut-off value of a target indicator allowing to classify an observation with respect to a given binary characteristics (Fawcett, 2005).

28. In particular, in the ROC-Is procedure the binary variable is represented by the suspect of under-reporting, while the classifier is represented by a composite indicator, which summarizes the information provided by the structural and economic indicators defined in the preceding subsection, and is built in four stages.

29. In the first stage, the binary variable representing the proxy of under-reporting ("suspect" of tax evasion) is defined based on the comparison, for each firm i, between the suitable performance indicator[5] and the average of the given stratum s. Therefore, each firm

[2] The ROC-Is procedure is not applied to the firms that have just one self-employed and are included in flat-rate tax regimes according to their industry, personal characteristics and territorial residence. For these productive units, an "adjusted" Franz method is still applied.

[3] Non-treatable firms are defined according to the following characteristics: (1) units belonging to Public Administration or operating in regulated markets; (2) units belonging to MNE groups, for which informative sources are lacking; (3) start-ups (active since less than one year); (4) units having M&A or bankruptcy procedures in the year; (5) units for which economic variables are determined by peculiar conditions (i.e. social cooperatives or units operating in buying and selling of real estate).

[4] Earnings Before Interest, Taxes, Depreciation and Amortization (EBITDA)

[5] For firms with less than 10 workers, the performance indicator is represented by the operative margin per self-employed ratio, while for firms with more than 10 workers, it is represented by the value added per worker ratio.

is classified as "suspect" (or not) according to whether its performance indicator is lower (or higher) than the stratum average.

30. In the second stage, for each stratum, a logit model is estimated having the proxy of "suspect" as dependent variable and the whole set of indicators as covariates. The results of the model allow to identify the five most informative indicators, which, in the third stage, are included in the factor analysis.

31. Finally, in the fourth stage, the three most informative factors are grouped to define, for each firm i in the given stratum, the composite indicator Z_i having the following form:

$$Z_i = \sum_j \omega_j F_{i,j} \qquad [1]$$

32. where ω_j are shares of explained variance for each factor, and the j-th factor is:

$$F_{i,j} = \sum_k \gamma_{k,j} \alpha_{i,k,j} \qquad [2]$$

33. 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.

2. Identification

34. The identification of under-reporting firms is based on the ROC analysis. In particular, starting from a logit model, ROC analysis identifies 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).

35. 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. The first is related to a statistical rather than subjective definition of the threshold clustering positive and negative observation.[6] The second is connected with the implicit correction that ROC analysis produces on the informative capability of the proxy variable.

36. The identification procedure is composed of three steps.

37. In the first, the following logit model is estimated:

$$Prob(proxy = 1) = \Lambda(\beta Z_i) \qquad [3]$$

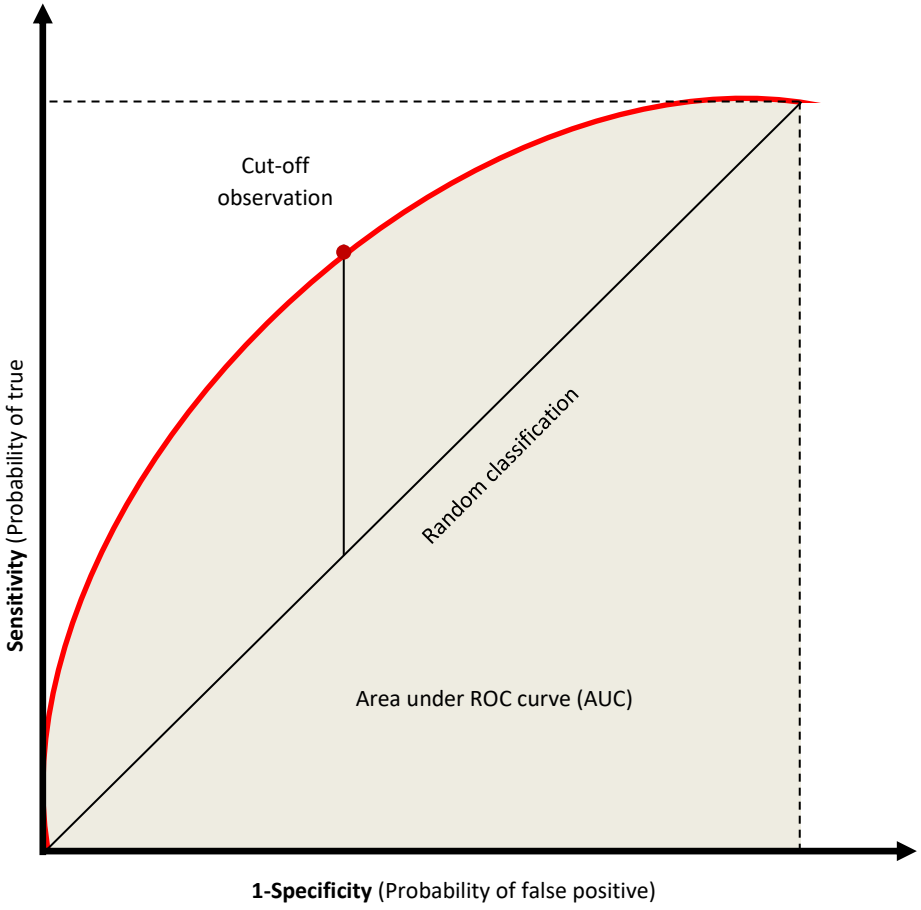
38. where Λ is the cumulative distribution of the logistic function, β is the estimated parameter and Z_i is the composite indicator.

39. The ROC curve in Figure 1 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. The ROC curve, therefore, also represents the trade-off between the possible errors in defining the clustering of observations (Kumar and Indrayan, 2011). Furthermore, the area under ROC curve (Area under ROC curve (AUC), the grey area in Figure 1) measures the extent to which

[6] Logit model provides for each observation a probability of belonging to the different groups defined by the binary variable. In this context, clustering of observation should depend on the threshold term of probability the researcher is willing to define. ROC analysis provides instead a method to statistically define this cut-off value based on an optimization procedure that takes into account the different weights the researcher is willing to assign to the different types of error (false negatives and false positives). See Fawcett (2005).

the covariate of the logit model is able to discriminate observations with respect to a random selection (represented by the 45° line).

Figure 1
The ROC curve



Source: Costa et al. (2019)

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

$$\widehat{Cut} = h * Sensitivity - (1 - h) * (1 - Specificity) \qquad [4]$$

41. 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 is set higher than 0.5, finding true positives is considered more relevant than incurring in false positives. If h is set lower than 0.5, avoiding false positives is considered more relevant than individuating true positives.

42. Given the value of h, Equation [4] permits to identify the cut-off observation along the ROC curve determined by the model in Equation [3].

43. 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. In particular, assuming $\bar{Z}$ as the value of the

composite indicator for the cut-off observation, other firms are classified as under-reporting if $Z_i < \bar{Z}$.

3. Adjustment

44. The adjustment of value added for under-reporting firms is obtained by exploiting the information provided by the identification stage. In particular, 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.

45. For each under-reporting unit in the given stratum, the following condition applies:

$$\sum_j \omega_j F_{i,j} < \bar{Z} \qquad [5]$$

46. where the first component is the value of the composite indicator for the i-th under-reporting firm and $\bar{Z}$ is the threshold value for the classification.

47. The adjustment is thus obtained using the following condition:

$$\tilde{\alpha}_{h,i} : \sum_j \omega_j F_{i,j} = \bar{Z} \qquad [6]$$

48. In particular, the composite indicator is incremented by leveraging on the value added-per-worker indicator ($\tilde{\alpha}_{h,i}$) as show in the following equation.

$$\tilde{\alpha}_{h,i} = \frac{\bar{Z} - \sum_j \omega_j \gamma_{j,-h} \alpha_{-h,i}}{\sum_j \omega_j \gamma_{j,h}} \qquad [7]$$

49. where, for the given stratum, $\tilde{\alpha}_{h,i}$ is the adjusted value added-per-worker indicator for the i-th under-reporting firm, $\bar{Z}$ is the threshold, ω_j are weights used to aggregate factors into the composite, $\gamma_{j,n}$ are the loadings representing the weight of each indicator α_k in the definition of the j-th factor (h represents the value added-per-worker indicator, while $-h$ represents other indicators).

50. Finally, the level of adjustment y_i can be obtained, for each under-reporting firm, in the given stratum, as:

$$y_i = \left(\tilde{\alpha}_{h,i} - \alpha_{h,i}\right) * N_i \qquad [8]$$

51. where $\tilde{\alpha}_{h,i}$ is the adjusted value added-per-worker indicator, $\alpha_{h,i}$ is the declared value added-per-worker indicator and N_i is the number of workers of the firm i.

52. For each under-reporting firm, the adjustment thus depends on both individual and general characteristics. Following Equation [7], the amount of adjustment is affected by three elements. First, the value of $\bar{Z}$, which is the threshold value in the given stratum. Second, the weight and the effect of the indicators other than the value added-per-worker (α_{-h}) in defining the value of the composite indicator (the second component of the numerator). Third, the weight and effect of the value added-per-worker indicator (α_h) in determining the value of the composite indicator (denominator).

53. In particular, the amount of adjustment 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. This direct relationship is modulated by the effects of the single indicators on the value of the composite. Indeed, the higher the "sensitivity" of the composite Z_i with respect the value added-per-worker (i.e. the denominator of Equation [7]), the lower will be the correction in the indicator which allows, ceteris paribus, to join the threshold for the given under-reporting firm. Symmetrically, the lower the "sensitivity" of composite Z_i with respect

to the other indicators (i.e. the second part of the numerator of Equation [7]), the higher will be the correction in the value added-per-worker which allows, ceteris paribus, to join the threshold for the given under-reporting firm.

III. Measuring VAT fraud

54. VAT fraud is strongly connected with under-reporting of value-added. Indeed, by hiding value added, firms also reduce the tax base for VAT.

55. Following the European Union Commission decision on the treatment for national accounts purposes of the non-collected VAT, VAT fraud has to be included in GDP and gross national income (GNI) according to the characteristics of the transactions generating it. In particular, two typologies of VAT fraud can be defined: (1) VAT fraud with complicity, which is generated in transactions in which parties consensually decide to hide the tax base; (2) VAT fraud without complicity, which is generated in transactions in which the seller hides the tax base, while the purchaser pays the related VAT.

56. In the compilation of National Accounts, the amount of VAT fraud with complicity has not to be considered, while GNI and GDP has to include the amount of VAT fraud without complicity. Indeed, while in the former case, the price of transaction does not include VAT, in the latter the price of transaction includes VAT, which is received by the seller and enters in his income (and, then, in the economic system).

57. In this context, in order to provide an estimate of VAT fraud, two main conditions have to be assured. The first relates to the capability of measuring the suitable tax base. The second is connected with the capability to distinguish between the two typologies of fraud (with and without complicity).

58. Considering the first issue, the ROC-Is method for measuring under-reporting provides a conceptually suitable assessment of the hidden tax base for VAT. In fact, in the preceding "Franz" procedure, the adjustment was connected with the total flow of resources entering in the firm, thus also including the amount of possible VAT fraud without complicity. In this case, a separate estimate of the tax base and VAT fraud was impossible. The new ROC-Is method, instead, adjusts for under-reporting considering only the value added at basic price, thus excluding possible amount coming from VAT fraud.

59. 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 VAT fraud with complicity and the one related to VAT fraud without complicity.

60. The procedure to estimate VAT fraud without complicity is composed of five stages, and relies on two main assumptions.

61. 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. There might be cases in which also in B2B transactions VAT fraud without complicity can occur.[7] There also might be cases in which complicity can occur also in B2C transactions.[8] The first issue is 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. The second issue is dealt with by identifying specific markets (industries) in which consumers and firms can have the possibility and the incentive to fraud tax authority consensually.

[7] Let assume that firm A buys from firm B a good. The transaction is rightly invoiced, but only firm A (the buyer) registers the invoice, while firm B (the seller) does not. In this case, firm A pays the whole amount (including VAT) to firm B, while the latter do not pay the relative VAT.

[8] Let assume Mr. White going to M.D. Black for a medical examination. The price for the medical examination is 100 euro excluding VAT. Mr. White can claim for an invoice of the price, which in this case will be 100 plus VAT, or he can pay only 100 without the invoice. If he decides to take the second solution then a VAT fraud with complicity emerges.

62. 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.

63. 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 amount of value added related to B2B and B2C transactions is determined and, then, applied to the given amount of adjustment for under-reporting:

$$y_{B2B,i} = \mu_{B2B,i} * y_i \qquad [9a]$$
$$y_{B2C,i} = \mu_{B2C,i} * y_i \qquad [9b]$$

64. where y_i is the amount of adjustment for firm i and μs are the share of declared value added coming from the different typologies of transactions.

65. In the second stage, implicit VAT rates, for each type of transaction (B2B and B2C) are calculated using the same dataset:

$$\rho_{B2B,i} = T_{B2B,i}/Y_{B2B,i} \qquad [10a]$$
$$\rho_{B2C,i} = T_{B2C,i}/Y_{B2C,i} \qquad [10b]$$

66. where Ts are the amount of VAT paid by the firm i for the relative typology of transaction, and Y_is are the amounts of declared value added by typology of transaction.

67. In the third stage, the amount of VAT fraud related to under-reporting of value added is then determined by multiplying the tax bases found in Equation 9A and 9B for the VAT rate defined in Equation 10A and 10B:

$$t_{B2B,i} = \rho_{B2B,i} * y_{B2B,i} \qquad [11a]$$
$$t_{B2C,i} = \rho_{B2C,i} * y_{B2C,i} \qquad [11b]$$

68. Following the first assumption, therefore, $t_{B2B,i}$ should represent the amount of VAT fraud with complicity, while $t_{B2C,i}$ should instead represent the amount of VAT fraud without complicity to be included in the GDP. However, as pointed out above, this assumption has to be relaxed in order to take into account possible peculiar cases.

69. The fourth stage is indeed aimed at correcting $t_{B2B,i}$ and $t_{B2C,i}$ in order to define the final measures of VAT fraud with and without complicity. Due to the lack of micro information with respect to these cases (cfr. Footnotes 6 and 7), these corrections are made at meso-level (by industry).

70. In particular, the first case (i.e. B2B transactions generating VAT fraud without complicity) is dealt with by correcting $t_{B2B,i}$ considering, for each industry s, the share of this type of transaction that are suspected of VAT fraud without complicity:

$$\bar{t}_{B2B,s} = \rho_{B2B,s} * y_{B2B,s} * \tau_{B2B,s} \qquad [12]$$

71. where τ_s measure the mismatch between inward and outward invoice in the s-th industry.

72. The second case (i.e. B2C transactions generating VAT fraud with complicity) is dealt with by assuming, for each industry, a share of B2C transactions that generates VAT fraud with complicity:

$$\bar{t}_{B2C,s} = \rho_{B2C,s} * y_{B2C,s} * \tau_{B2C,s} \qquad [13]$$

73. Finally, in the fifth stage, Equations [12] and [13] are used to define the amount of VAT fraud with and without complicity, by correcting the results coming from Equations [11a] and [11b]:

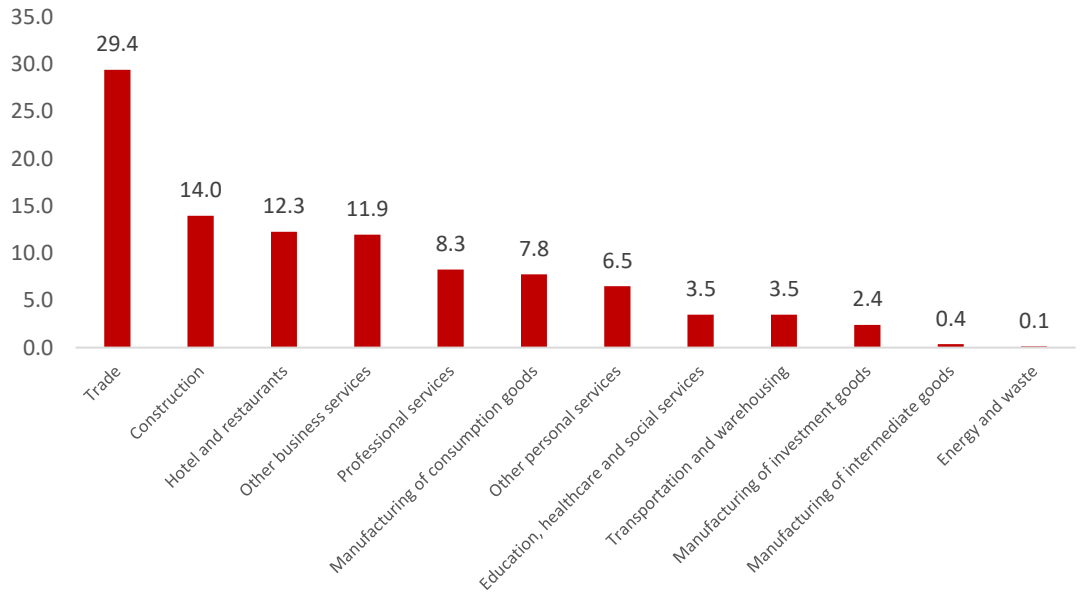
$$t_C = \sum_s (\sum_i t_{B2B,i}) - \bar{t}_{B2B,s} + \bar{t}_{B2C,s} \qquad [14a]$$
$$t_N = \sum_s (\sum_i t_{B2C,i}) - \bar{t}_{B2C,s} + \bar{t}_{B2B,s} \qquad [14b]$$

74. where t_C is the final amount of VAT fraud with complicity and t_N is the final amount of VAT fraud without complicity.

IV. Results

75. 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.[9] 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.

Figure 2
Share of under-reporting units by industry with respect to total under-reporting units



76. Considering the whole sub-population, Figure 2 shows the weight of each industry in the total number of under-reporting firms. In this case, 29.4% of under-reporting firms are included in wholesale and retail trade, while construction, and hotel and restaurants accounts for, respectively, 14.0% and 12.3%. On the other hand, under-reporting firms in the manufacturing of investment and intermediate goods, or in the energy and waste industry account for less than 3% of the total.

77. These results are relevant in determining where in the business system under-reporting is more widespread. On the other hand, also a relative measure of the impact of under-reporting by industry can help deepening the analysis. In this context, Figure 3 shows the share of under-reporting firms by industry.

78. Comparing industries with the total average, results stress a relevant sectoral heterogeneity. Indeed, on the one hand, hotel and restaurants and other personal services show a share of under-reporting firms out of the total of firms with less than 100 employees over the 70%, while, on the other hand, industries such as intermediate goods, professional services and education, healthcare and social services have shares under the 40%.

[9] As noted above, the under-reporting of a part of firms are still estimated using a "modified" Franz-method.

Figure 3
Share of under-reporting units by industry

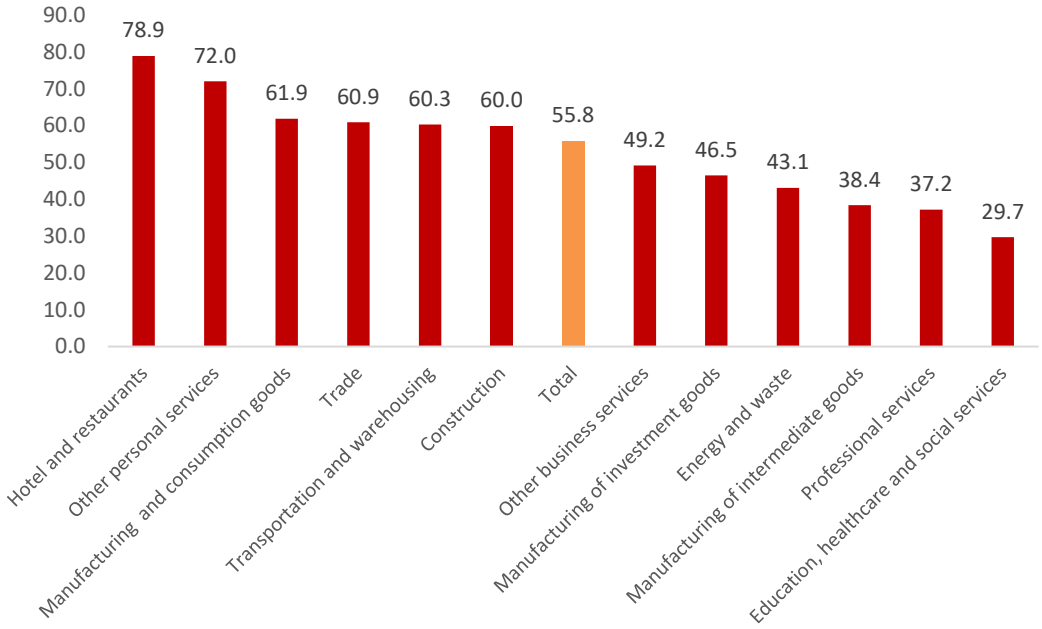
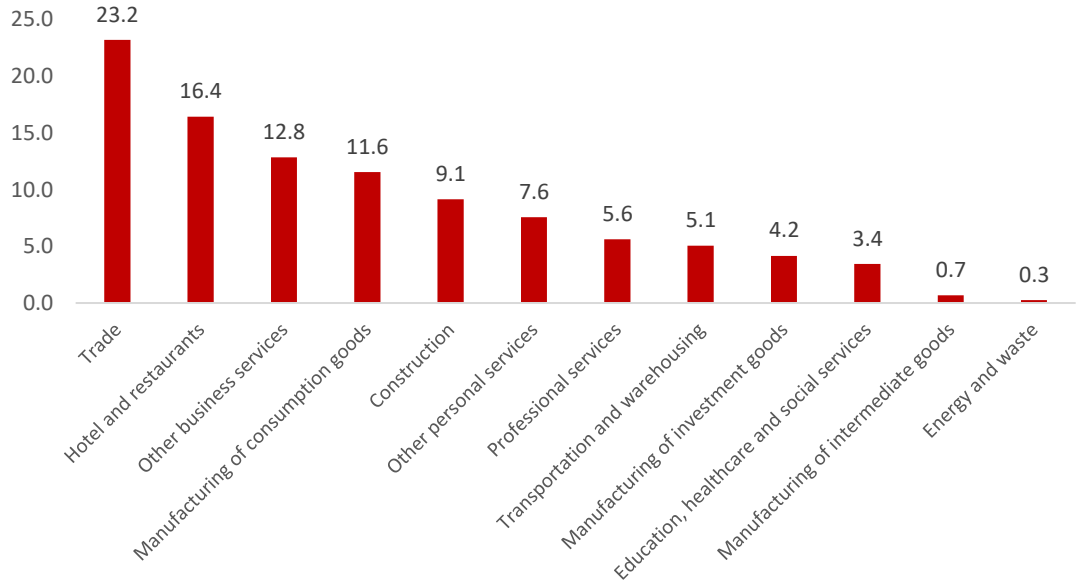
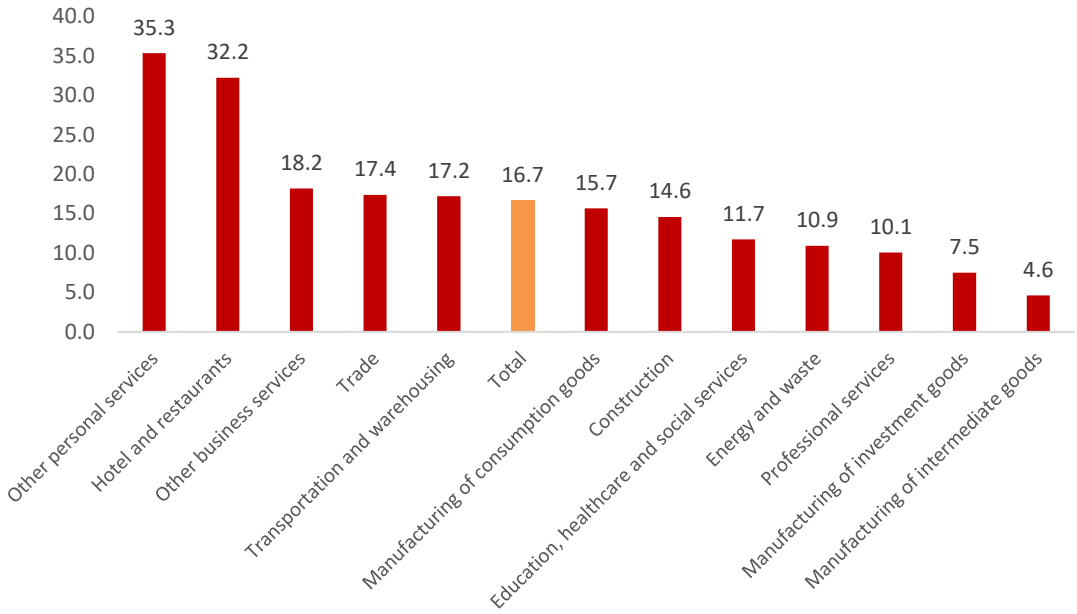


Figure 4
Share of adjusted value added by industry with respect to total adjustment



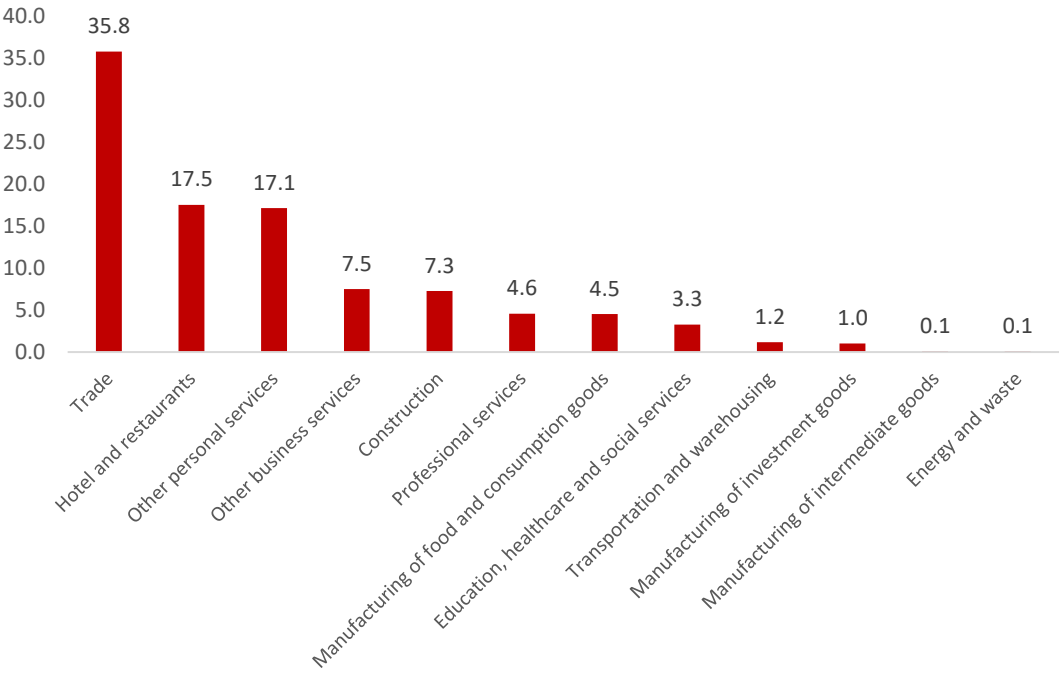
79. Shifting from the identification of under-reporting firms to adjustment of her value added, Figure 4 shows that the magnitude of under-reporting in terms of value added is strongly concentrated. Indeed, the five industries with the highest value of adjustment (i.e. trade, hotel and restaurants, other business services, production of food and consumption goods, and construction) explain about the 75% of the total hidden value added.

Figure 5
Share of adjustment in total adjusted value added by industry



80. Also in this case, however, a relative measure of the impact of under-reporting by industry allows for deepening the analysis. Indeed, the share of adjustment with respect to total value added is highest in other personal services (35.3%), and hotel and restaurants (32.2), while in other industries such as trade and construction is much lower (respectively 17.4% and 14.6%).

Figure 6
Share of VAT fraud without complicity by industry with respect to total VAT fraud without complicity



81. VAT fraud without complicity related to under-reporting is about 4.5 billion euros, and represents about the 80% of the total.[10] Trade generates 29.4% of the total fraud, followed by construction (14.0%), hotel and restaurants (12.3%) and other business services (11.9%).

V. Conclusion

82. The aim of this work was presenting the new procedure developed by ISTAT to estimate under-reporting of value added.

83. The new ROC-Is method has been conceived with the goal of overcoming the set of limitations of the old "Franz" procedure.

84. In particular, the conceptual reliability of the ROC-Is method is independent from the size of firms, results are not forcedly anti-cyclical and the procedure exploits a large part of the huge information contained in the Frame-SBS archive. Furthermore, the procedure is coherent with the possibility of separately measuring also VAT fraud without complicity, which has to be included in GDP and GNI.

85. The ROC-Is method is based on the application of the ROC analysis to a composite indicator representing the economic behavior of firms and to a classification variable representing the "suspect" of under-reporting. This model allows for both individuating under-reporting firms and adjusting the value added.

86. 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, which is in turn dependent on the distribution of the economic behavior of firms (represented by the composite indicator). In this respect, this method permits to classify production units taking into account both their peculiar characteristics and the general trend and structure of the industry, avoiding any subjective evaluation.

87. Considering VAT fraud, the possibility to have a micro database of under-reporting and the conceptual coherence of the ROC-Is method permits to measure the phenomenon separately from under-reporting. This 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.

[10] About 20% of VAT fraud without complicity relates to the value added generated using unregistered workforce, which is out of the scope of this work.

References

Costa, S., Sallusti, F., Vicarelli, C., Zurlo, D. (2019). Over the ROC methodology: Productivity, economic size and firms' export thresholds. *Review of International Economics*, 00: 1–26. doi: 10.1111/roie.12405

Fawcett, T. (2005). An introduction to ROC analysis. *Pattern Recognition Letters*, 27: 861-874. https://doi.org/10.1016/j.patrec.2005.10.010.

Franz, A. (1983), Wie groß ist die "schwarze" Wirtschaft? *Mitteilungsblatt der Österreichischen Statistischen Gesellschaft*, 49: 1–6.

ISTAT (2018). L'economia non osservata nei conti nazionali. ISTAT, Rome.

Kumar, R., Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics*, 48(4): 277–287. https ://doi.org/10.1007/s13312-011-0055-4

Luzi, O., Monducci, R. (2016). The new statistical register 'Frame SBS': overview and perspectives. *Rivista di statistica ufficiale*, 1: 5-14.

OECD (2002). Measuring non-observed economy: A handbook. OECD, Paris.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3: 32–35. https://doi.org/10.1002/1097-0142(1950)