TECHNICAL ASSISTANCE REPORT

REPUBLIC OF ARMENIA

Corporate Income Tax Gap Estimation Based on Operational Audits

MAY 2024

Prepared By

Soren Pedersen, Tobias Gabel Christiansen

CCAMTAC

MEMBERS

Armenia, Azerbaijan, Georgia, Kazakhstan, Kyrgyz Republic, Mongolia, Tajikistan, Turkmenistan, Uzbekistan

PARTNERS

Switzerland, Russia, China, Korea, United States, Asian Development Bank, Poland

DISCLAIMER

The contents of this report constitute technical advice provided by the staff of the International Monetary Fund (IMF) to the authorities of the Armenia (the "TA recipient") in response to their request for technical assistance. This report (in whole or in part) or summaries thereof may be disclosed by the IMF to IMF Executive Directors and members of their staff, as well as to other agencies or instrumentalities of the TA recipient, and upon their request, to World Bank staff, and other technical assistance providers and donors with legitimate interest, including members of the Steering Committee of CCAMTAC, unless the TA recipient specifically objects to such disclosure (see Operational Guidelines for the Dissemination of Technical Assistance Information). Publication or Disclosure of this report (in whole or in part) or summaries thereof to parties outside the IMF other than agencies or instrumentalities of the TA recipient, World Bank staff, other technical assistance providers and donors with legitimate interest, including members of the Steering Committee of CCAMTAC, shall require the explicit consent of the TA recipient and the IMF's Fiscal Affairs Department.

MEMBERS

Armenia	Azerbaijan	Georgia	Kazakhstan	Kyrgyz Republic
Mongolia	Tajikistan	Turkmenistan	Uzbekistan	

DEVELOPMENT PARTNERS (DONORS)

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Economic Affairs SECO

Ministry of Economy and Finance

EUROPEAN UNION

ADB

Ministry of Foreign Affairs Republic of Poland

Contents

Abbreviations and Acronyms

AMD	Armenian Dram is the currency of Armenia
CCAMTAC	The Caucasus, Central Asia, and Mongolia Regional Capacity Development Center
CD	Capacity Development
CIT	Corporate Income Tax
CRM	Compliance Risk Management
FAD	IMF Fiscal Affairs Department
GDP	Gross Domestic Product
IMF	International Monetary Fund
IMR	Inverse Mills Ratio
MLM	Machine Learning Model
NACE	Nomenclature statistique des Activites economiques dans la Communaute Europeenne
PIT	Personal Income Tax
SRC	Armenia State Revenue Committee
STX	IMF Short Term Expert
TIN	Taxpayer Identification Number
VAT	Value Added Tax

Preface

In response to a request from the Armenia State Revenue Committee (SRC), a capacity development (CD) mission team comprising Mr. Soren Pedersen (FAD) and Mr. Tobias Gabel Christiansen (FAD short-term expert), visited Yerevan, Armenia during the period March 11 – 22, 2024. The purpose of this mission, financed by CCAMTAC, was to estimate the Corporate Income Tax (CIT) Gap based on operational audits as well as providing SRC with results from a machine learning model (MLM) to select cases for future audits.

Prior to the visit to Yerevan, a “data requirement folder” was sent to SRC and SRC had provided most of the required data. Several productive workshops and meetings were held with staff of SRC to ensure an understanding of the data provided by SRC for the estimations and clarify other technical issues. The report’s findings are based on the data provided by the SRC.

The IMF team expresses its sincere appreciation to SRC for the excellent cooperation, fruitful workshops and meetings, and the excellent support provided both before and during this mission. The team particularly acknowledges the excellent support provided by Mr. Arsen Sarikyan, Head of Development and Administration Strategy Programs Department; Mr. Mkhitar Ayvazyan, Deputy Head of Development and Administration Strategy Programs Department; Mr. Martin Sandoyan, Head of Revenue Assessment and Analysis Division; and Grigor Hakobyan, Chief Specialist.

This report represents the final version of the draft report that was submitted to Mr. Ashot H. Muradyan, Deputy Chairman and Mr. Arsen Sarikyan, Head of Development and Administration Strategy Programs Department, on March 21, 2024. It consists of an Executive Summary and the following sections: (I) Introduction (II) Key Findings; and (III) Next Steps.

Executive Summary

This mission, financed by CCAMTAC, estimated the corporate income tax (CIT) Gap in Armenia based on a bottom-up approach using data from the State Revenue Committee's (SRC) operational audits.[1] The CIT Gap is estimated as a three-year average at 35.2 percent (or 26.4 percent)[2] of potential CIT liability for the income years 2020, 2021, and 2022.

For the years 2020, 2021, and 2022, the CIT Gap is estimated at 38.8, 35.2, and 32.8 percent of potential CIT liability respectively.[3] Excluding audit adjustments with no immediate impact on revenue (where the audit adjustment resulted in reducing a loss or carry forward of losses), the overall gap is 26.4 percent. Split by year it is 29.2, 26.4, and 24.6 percent, in 2020, 2021, and 2022, respectively.

Average CIT Gap was measured at 1.4 percent of GDP. For the years 2020, 2021, and 2022, the CIT Gap is 1.3, 1.4, and 1.5 percent of GDP, respectively. Excluding audit adjustments with no immediate impact on revenue the corresponding the numbers are1.0 percent overall, 0.9 percent for 2020, 0.9 percent for 2021, and 1.0 percent for 2022.

In absolute terms, the estimated CIT Gap overall is 93.9 billion AMD, 83.1 billion AMD in 2020, 90.3 billion AMD in 2021, and 108.5 billion AMD in 2022. Excluding audit adjustments with no immediate impact on revenue the overall CIT gap is 62.0 billion AMD. For 2020, 2021, and 2022, the CIT gap is 54.0 billion AMD, 59.6 billion AMD, and 72.4 billion AMD.

The machine learning model (MLM) deployed has the potential to enhance audit results in Armenia and optimize resource allocation on CIT audits. A MLM was also developed, which can be used to identify high risk cases for new audits.[4] The MLM predicted that auditing the top 10 percent of corporations with the highest risk is likely to raise revenue by three times compared to SRC's existing risk model. The MLM also results in higher average CIT corrections compared to the current SRC strategy when auditing taxpayers other than those in the top 10 percent.

As next step it was agreed to train SRC analysts in applying the MLM on 2023 CIT returns to select cases for audits. It was also agreed as a next step to test the effect of the MLM versus the current selection strategy.

The outcome from this exercise can be used in SRC's Compliance Risk Management (CRM) work. If the test of the MLM is shown to be more effective in terms of audit outcomes, the results from the MLM should be utilized in SRC's future CRM efforts to select high-risk cases for audit. This would enhance the efficiency of future audits and has the potential to increase the revenue collected from audits.

---

[1] This analysis measures the CIT compliance gap, i.e., the policy gap is excluded. For simplicity, "gap" and "compliance gap" will be used indistinguishably in this report.

[2] Excluding audit adjustments with no immediate impact on revenue (due to tax losses before and after audit). The estimates do not account for undetected noncompliance which could lead to underestimation of the CIT gap.

[3] Potential CIT liability is defined as self-reported CIT plus the estimated CIT gap.

[4] See paragraph 3 and 4 for a short description of the model.

Recommendations

Test cases selected from MLM		Due data
1	Test the effectiveness of the MLM by selecting 50 percent of cases for income year 2023 using the traditional risk-based model and 50 percent of cases using the MLM.	April 2025
1.1	As a prerequisite for this, update the already delivered data with new data for income year 2023 in exactly the same way to select cases for audit for the income year 2023.	May 2024
1.2	Seek IMF capacity development (CD) to train staff and help select cases for income year 2023 using the developed MLM.	May 2024
1.3	Assess the results of the two risk selection methods and decide which model should be used for selecting audits for income year 2024.	April 2025
1.4	Examine the reasons for non-compliance more closely in high-risk sectors such as the mining sector to understand if it is due to deliberate evasion, lack of knowledge of the tax law or complicated legislation. This will give valuable information for a compliance improvement plan.	April 2025
Data analytics		
2	Appoint a team of 2-3 people to engage in data analytics so that SRC can run the developed models itself.	June 2024
3	The appointed team should invest around 80 percent of their time on data analytics and learn the necessary programming tools to be able to carry out CIT Gap analysis and to maintain the MLM independently.	December 2024

I. Introduction

1. This mission, financed by the CCAMTAC, estimated the CIT Gap based on a bottom-up approach using data from operational audits. In Armenia, it is prohibited by law to carry out random audits which is the ideal tool to measure the CIT Gap based on a bottom-up approach. Due to the lack of random audits, SRC requested the CIT Gap be measured using operational audits.

2. Using operational audits to measure the CIT Gap needs to account for the non-random selection of audited corporations. Without correction for cases not selected randomly, the CIT Gap would be overestimated. This bias, known as "Sample Selection Bias," arises from the non-random selection of cases. The method devised by James J. Heckman is utilized to correct for Sample Selection Bias.[5]

3. The CIT gap was thus estimated using the Heckman Sample Selection model. The Heckman method is a two-stage procedure. In the first stage, it estimates the probability that a company is selected for audit. This is done using a probit model[6] and 25 separate risk scores used by the SRC to target CIT audits. The second stage models the audit outcome (i.e., tax uncovered) using company characteristics (i.e., lines from the CIT return, sector, number of employees etc.) and a regressor that accounts for the selection process (see Annex I for more details).

4. The Heckman two-step estimator's suitability in estimating the CIT gap depends on how audits are conducted. If audits focus narrowly on specific parts of a business, they may miss undisclosed taxes, leading to an underestimated CIT gap. Additionally, when audits target specific companies, such as a particular sector or type of firm, it becomes challenging to obtain a reliable estimate of the tax. This difficulty arises from the increased extrapolation required between audited companies and the rest of the population. Finally, precautions must be taken since a fraction of noncompliance could be undetectable, even under the best efforts of auditing. This possibility could result in an underestimation of the CIT gap

5. CIT Gap estimates were obtained for three consecutive years. These estimates cover the latest available data for the income years 2020, 2021, and 2022. Due to the limited annual number of audits, the yearly estimates are more uncertain.

6. In addition to CIT Gap estimation, a MLM was developed. It was possible to develop a MLM to select cases for future audits, as the data used for both the Heckman method and developing the MLM is the same.

7. Even though the MLM looks promising, it must be tested on new tax declarations for the income year 2023. Before prioritizing future audits based on the new MLM, the new model should be tested against the SRC’s existing model for selecting audits on CIT returns filed for the income year 2023.

[5] James J. Heckman (1979). “Sample Selection Bias as a Specification Error”. Econometrica. vol. 47(1), pp. 153-161.

[6] In a probit model the outcome is binary (0 or 1) – in this case whether a company has been audited or not.

The test should compare the tax uncovered from audits according to each model, as this is the main criterion for success according to the SRC.

8. The outcome from this exercise can be used in SRC's Compliance Risk Management (CRM) work. If the test of the MLM confirms initial results that the MLM is more effective in terms of audit outcomes, the results from the MLM should be utilized in SRC's future CRM efforts to select high-risk cases for audit. This would enhance the efficiency of future audits and has the potential to increase the revenue collected from audits.

II. Key Findings

A. CIT Gap results

9. **The CIT Gap is estimated to be 35.2 percent of potential CIT liability, on average, for the income years 2020, 2021, and 2022.** The average CIT Gap is estimated based on operational audits conducted for the income years 2020, 2021, and 2022. The result is based on a total of 4,432 comprehensive audits covering all aspects of the company, carried out by the SRC over the three-year period. The results have been adjusted to account for the non-random selection of operational audits, addressing what is known as "sample selection bias."

10. **The CIT Gap is estimated to be 38.8 percent of potential CIT liability in 2020, 35.2 in 2021, and 32.8 in 2022 (Table 1, Panel A)** The number of audits conducted in the three years were 1,886 in 2020, 1,483 in 2021, and 865 in 2022. Due to the lower number of audits conducted each year, the results for each year are more uncertain compared to the estimated average CIT Gap based on pooled data.

11. **Overall, the CIT Gap is 1.4 percent of GDP in 2020-2022 and 1.3, 1.4, and 1.5 percent of GDP in 2020, 2021, 2022 respectively.** The results are summarized in Table 1, Panel A.

Table 1. CIT Gap estimates based on operational audits

	Panel A: Using all audit adjustments			
	Overall	**2020**	**2021**	**2022**
CIT gap	93.9 billion AMD[1)]	83.1 billion AMD[1)]	90.3 billion AMD[1)]	108.5 billion AMD[1)]
Percent of potential CIT	35.2	38.8	35.2	32.8
Percent of GDP	1.4	1.3	1.4	1.5
	Panel B: Excluding audit adjustments with no immediate effect on revenue			
	Overall	**2020**	**2021**	**2022**
CIT gap	62.0 billion AMD[1)]	54.0 billion AMD[1)]	59.6 billion AMD[1)]	72.4 billion AMD[1)]
Percent of potential CIT	26.4	29.2	26.4	24.6
Percent of GDP[2)]	0.9	0.9	0.9	1.0

Source: IMF calculations based on data from SRC.
Note: 1) In 2020-prices. 0.1% of audit results were trimmed in top and bottom for each year. Re-audits are discarded. Only corporations with a reported CIT return. Potential CIT liability is defined as self-reported CIT plus the estimated CIT gap. Panel A presents estimates of the CIT gap using all audit adjustments. Panel B excludes audit adjustments with no immediate effect on revenue (i.e., companies with a net-loss before and after audit). 2) GDP is also in 2020 prices in the calculation of the CIT gap in percent of GDP.

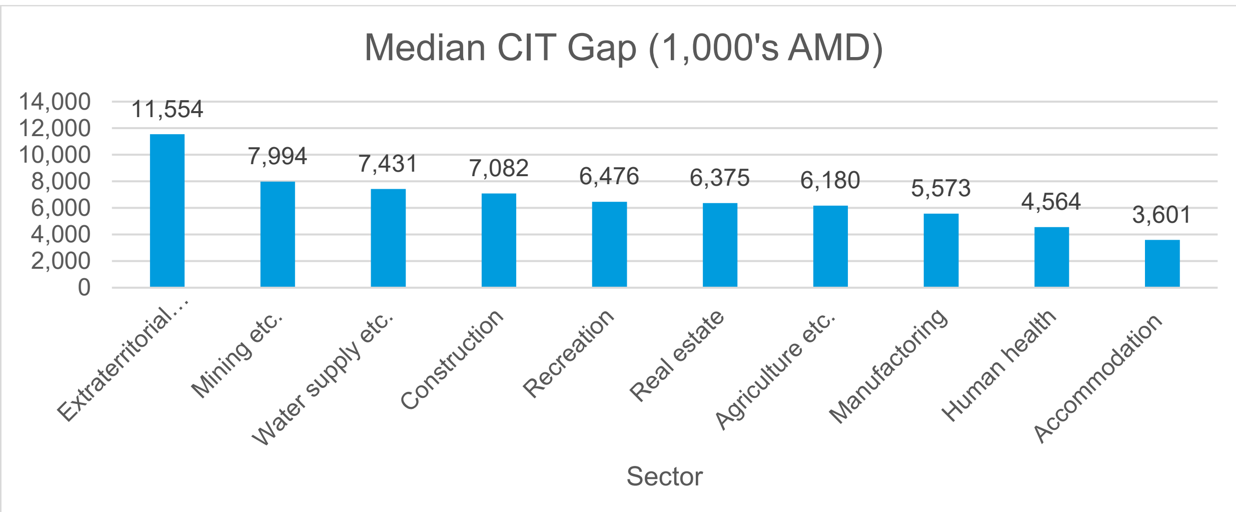
12. **If audit adjustments with no immediate effect on revenue are excluded, the CIT Gap is 26.4 percent of potential CIT or 0.9 percent of GDP over the years 2020-2022, compared to 35.2 percent of potential CIT or 1.4 percent of GDP (Table 1, Panel B).** The immediate tax liability of companies with a net loss before and after an audit remains unchanged, and whether to include it in the tax gap or not is

subject to debate. On one hand, the audit adjustment has no immediate effect on revenues. On the other hand, if the losses of a company are reduced following an audit, the losses that the company can carry forward to future years are reduced, leading to higher future tax revenues (except in cases where the company goes bankrupt, is liquidated, or continues to evade taxes etc.).[7]

13. **Excluding the largest corporations lowers the nominal value of the CIT gap, while the CIT gap in terms of potential tax remains similar (Table 1B in Annex**). Removing the top 0.1 percent largest companies, measured by turnover, results in a CIT gap of 64.9 billion AMD or 34.2 percent of potential CIT (47.5 billion AMD or 27.9 percent of potential CIT excluding audit adjustments with no immediate effect on revenue). Due to the limited number large companies, the associated tax gap entails more uncertainty. Excluding these companies can reduce 'noise' at the cost of reducing the population for which the tax gap is estimated.[8]

14. **The median CIT Gap is highest in the sector with "Extraterritorial" activities (Figure 1).[9]** However, it should be noted that this sector is small, so its contribution to the total CIT Gap is limited. As seen in Figure 1, the second-highest median CIT Gap is in the mining sector. The lowest median CIT Gap among the top-10 sectors is found in the accommodation sector.

Figure 1. Median CIT Gap (1,000's AMD) by top-10 sectors[10]



Source: IMF calculations based on data from SRC.

15. **It is important to examine the reasons for non-compliance more closely in high-risk sectors such as the mining sector.** It will give valuable information for a compliance improvement plan

---

[7] In Denmark, for example, all corrections are included in the estimation of the tax gap, including those with no immediate impact on revenue (source: https://www.ft.dk/samling/20171/almdel/SAU/bilag/92/1839723.pdf, available in Danish only). On the other hand, the Inland Revenue Service (IRS) in the United States (US) does not include corrections in their tax gap estimation it the audit adjustment results in no payment – see page 29 in [Publication 5784 (10-2022) (irs.gov)](https://www.irs.gov).

[8] The IRS CIT gap methodology excludes Large Businesses from the scope of the Heckman methodology.

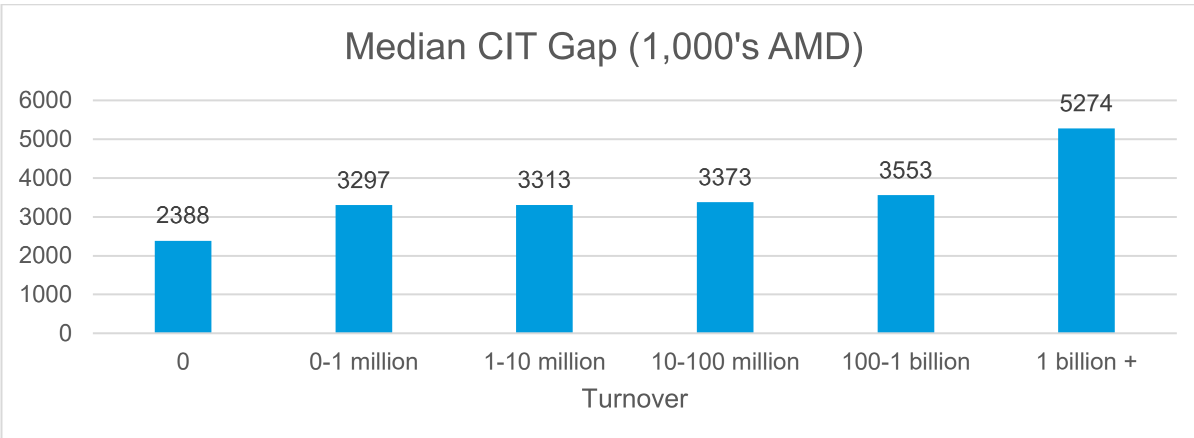
[9] Sector classification follows NACE coding (Nomenclature statistique des Activites Economiques'-Statistical classification of economic activities in the European Community).

[10] This sector is considered the primary sector for a corporation if they operate in more than one sector. Sector codes are self-reported by corporations. The chart only displays the top 10 sectors due to space limitations.

to understand if non-compliance is due to deliberate evasion, lack of knowledge of the tax law or complicated legislation.

16. The median CIT gap rises with turnover (Figure 2). The median CIT Gap is lowest among corporations with turnover under one million AMD. The median CIT Gap is then quite stable for corporations with turnover from one million AMD to one billion AMD. The highest median CIT Gap is among corporations with over one billion AMD in turnover.

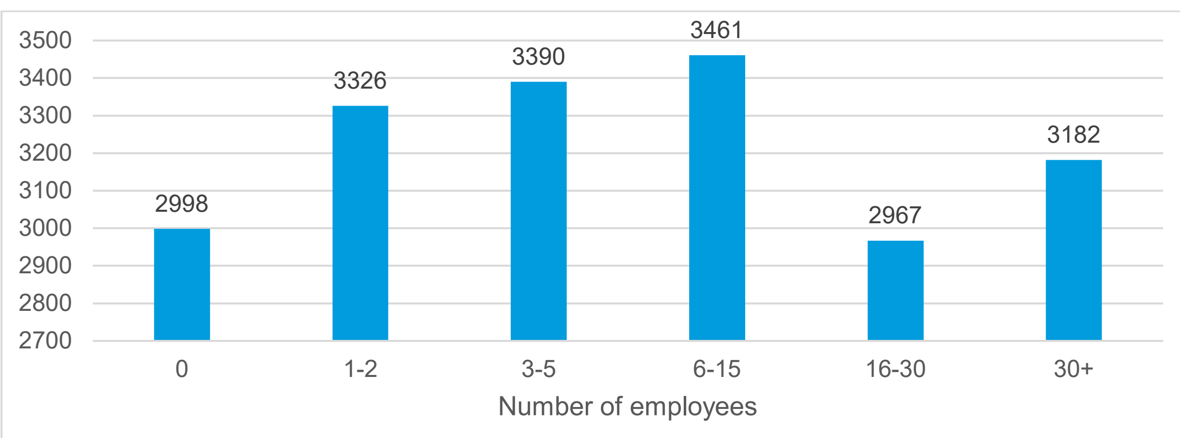
Figure 2. Median CIT Gap (1,000’s AMD) by turnover



Source: IMF calculations based on data from SRC.

17. There is no clear correlation between number of employees in corporations and the median CIT Gap (Figure 3). Number of employees in corporations serves as another metric to measure the size of the company. However, as depicted in Figure 3, there is no distinct correlation between the median CIT Gap and the number of employees in corporations. The highest median CIT Gap is observed in corporations with 6-15 employees.

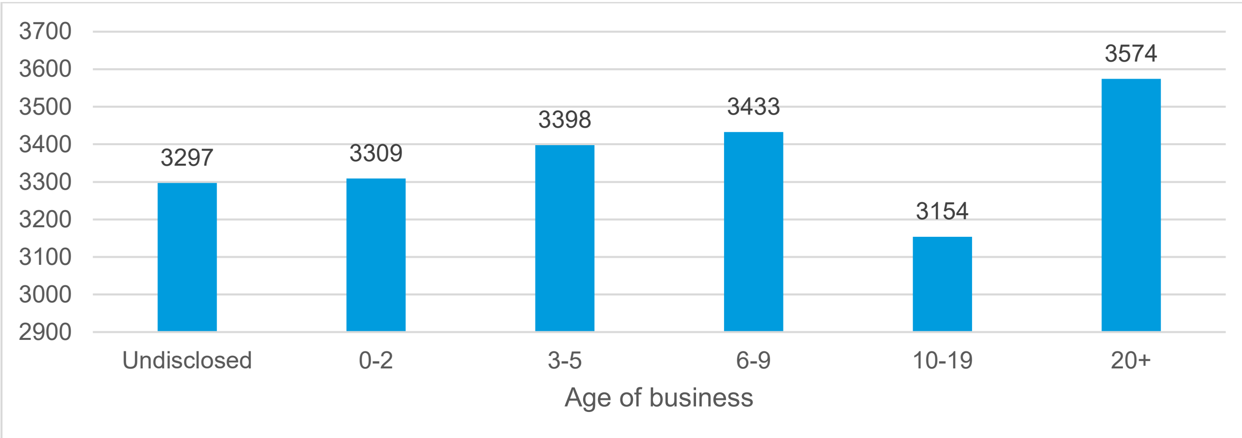
Figure 3. Median CIT Gap (1,000’s AMD) by number of employees



Source: IMF calculations based on data from SRC.

18. There is little variation between the median CIT Gap and age of corporations (Figure 4). The highest median CIT Gap is found among corporations that are more than 20 years old. But the median CIT Gap in this group is only marginally higher than corporations that are 3-5 and 6-9 years old.

Figure 4. Median CIT Gap (1,000's AMD) by age of business



Source: IMF calculations based on data from SRC.

B. Operational audits

19. The 'hit rate' based on operational audits was 44.3 percent in 2020, 44.2 percent in 2021, and 32.6 percent in 2022 (Table 2).[11] The hit rate based on operational audits was almost the same in 2020 and 2021 of 44 percent. The hit rate seems to decline to around 33 percent. in 2022. However, audits for the income year 2022 are not yet completed so the final hit rate for 2022 will likely be higher typically because difficult/highest evasion cases take more time to complete.

Table 2. Hit rate, average correction for income years 2020, 2021 and 2022

	2020	2021	2022
Hit rate	44.3	44.2	32.6
Average correction	3.6 m. AMD[1)]	4.7 m. AMD[1)]	3.2 m. AMD[1) 2)]
Number of audits	1,886	1,483	865

Source: IMF calculations based on data from SRC.

Notes: [1)] In 2020-prices. [2)] Lates completion date in data was Dec. 29, 2023. 0.1 pct. of audit results were trimmed in top and bottom for each year. Re-audits are discarded. Only corporations with a reported CIT return.

20. When correcting for the fact that there are still approximately 3.5 months left for SRC to audit tax declarations for the income year 2022, there seems to be a small decrease in the hit rate from 2020 to 2022. The most recent completed audit for the income year 2022 was on December 29, 2023. Focusing solely on audits for the income year 2020 completed before December 29, 2021, reveals a hit rate of 41.4 percent. Similarly, audits for the income year 2021 completed before December 29, 2022, show a hit rate of 35.0 percent. This indicates a decreasing hit rate over time. Possible reasons for

[11] The hit rate is defined as number of companies with audit adjustment divided by number of companies audited.

this trend could include higher compliance levels or updates to the strategy used by the SRC to select audits.

21. The lack of random audits makes it difficult to say whether the decrease in the hit rate is due to changes in SRC's risk scores or increased compliance among taxpayers. SRC reassesses its risk scores annually. In the income year 2020, SRC utilized 19 different risk scores, 20 in 2021, and 25 in 2022. Furthermore, each risk is evaluated annually and may undergo slight alterations.

22. Average correction per audit has risen from 2020 to 2021 (Table 2). It is premature to determine the average correction per audit for the income year 2022, as SRC still has approximately 3.5 months to audit cases related to 2022.

C. Machine learning model

23. Using machine learning to identify high-risk cases for audits leads to a threefold increase in the CIT uncovered from audit compared to the current strategy used by the SRC. Prioritizing audits based on the top 10 percent highest risk scores from the MLM results in an average correction of 15.7 million AMD, (see Table 3). This is three times higher than allocating audits based on the top 10 percent highest risk scores according to the SRC's current method, which results in an average correction of 5.2 million AMD. The MLM continues to result in higher average CIT corrections compared to the SRC method when auditing beyond the top 10 percent.

24. The MLM has a slightly lower hit rate compared to the SRC strategy. The hit rates when auditing the top 10 percent and top 20 percent according to the MLM are 43.7 percent and 43.9 percent, respectively. These rates are lower than the hit rates obtained when auditing the top 10 percent and top 20 percent according to the SRC scores, which are 44.9 percent and 51.4 percent, respectively. The hit rates become nearly identical when auditing more than the top 20 percent.[12]

25. The MLM was developed to identify cases where the CIT uncovered from audit is high. Based on the outcomes from operational audits and information from CIT returns filed by companies, the model is built to identify large audit adjustments. Targeting the size of the audit adjustments leads to large average audit adjustments but lower hit rates compared to models that target a high hit rate. The decision to target large audit adjustments was made by the SRC.

26. The model was trained on data from 2020 and 2021 and tested on data from 2022. To reflect that historic data is used to guide future decision-making, the model was trained on data from operational audits of CIT returns for the tax years 2020 and 2021 and evaluated on operational audits of CIT returns for the tax year 2022.

27. The MLM should be tested on CIT returns for the tax year 2023 to obtain an accurate measurement of its performance. The MLM was tested on operational audits for the tax year 2022. This data is not representative of companies in Armenia, as they were selected using risk scores developed by the SRC. To obtain an accurate estimate of the performance of the MLM, it should be tested by allowing it to freely choose 2023 CIT returns for audit.

---

[12] When applying a MLM, the analyst must decide to target either the hit rate or revenue. A MLM is often better at targeting the hit rate but that is on the cost of lower revenue. Popular speaking it is more difficult for the MLM to find rare incidents of high audit adjustments. On the other hand, it is easier for the MLM to find cases with audit adjustments (no matter of the size of the audit adjustment) because they are more common than cases with high audit adjustments. In other words, there is a trade-off between targeting the hit-rate or revenue. In the case of Armenia, the SRC wanted to focus on revenue (see next paragraph).

Table 3. Comparison of MLM and the current SRC strategy

	MLM		Current SRC strategy		Overlap
	Average correction, in million AMD	Hit rate, in percent	Average correction, in million AMD	Hit rate, in percent	Percent
Top 10 percent	15.7	43.7	5.2	44.9	37.1
Top 20 percent	10.7	43.9	7.0	51.4	50.9
Top 30 percent	8.3	52.3	6.2	54.0	59.3
Top 40 percent	7.6	55.5	6.5	55.3	72.3
Top 50 percent	6.3	56.9	5.9	53.6	86.0
All	3.2	32.6	3.2	32.6	100.0

Source: IMF calculations based on data from SRC. Note: Monetary values in 2020-prices.
Overlap measures the overlap in the selected companies according to each strategy.

III. Next Steps

A. Set up a team of data analysts

28. It is recommended that SRC appoints 2-3 data analysts with responsibility to compile future CIT Gap estimations. To sustain the CIT Gap model, it is crucial that SRC invests in data analytics. It is recommended that at least 2-3 individuals be trained in data analytics. This will enable SRC to independently conduct CIT Gap analysis based on operational audits developed by this mission.[13]

29. Data analysts should also be able to update the MLM to leverage its benefits. The MLM projects that audit results can be three times higher for highest 10 percent of risk cases compared to the existing risk model used by SRC. The MLM also results in higher average CIT corrections compared to the SRC method when auditing beyond the top 10 percent. Therefore, it is important that SRC can maintain and develop the MLM in the future.

30. The appointed team should invest around 80 percent of their time on data analytics. It is important to invest sufficient time to be able to perform good data analytics. The team should learn the necessary programming tools to be able to carry out CIT Gap analysis and to maintain the MLM independently.

B. Test the machine learning model

31. The MLM should undergo testing on CIT returns for the income year 2023. As illustrated earlier, the MLM was trained on tax returns for the income years 2020 and 2021, and subsequently tested on 2022 income year data. This approach reflects the ideal method of developing and testing the model, as it simulates real-world conditions. However, it is essential to conduct testing on data from the income year 2023 to ascertain whether the model is also more efficient in an "out of sample" context compared to the existing model used by SRC.

32. It is recommended that 50 percent of audits for the income year 2023 is selected using the MLM. This recommendation stems from the test results mentioned earlier, which indicated that utilizing the MLM for the top 10 percent highest risk scores led to a threefold increase in audit results compared to SRC's current strategy. Furthermore, the MLM consistently produces higher average CIT corrections compared to the SRC method.

33. Compare the results of the two risk selection methods and decide which model should be used for selecting audits for income year 2024. Based on the results from comparing the two selection

[13] The model is based on the XGBoost algorithm. The model implemented in R using the caret package (Kuhn, 2008), which implements the tree boosting model (XGBoost) as presented in Chen and Guestrin (2016). The hyperparameters are tuned using 5-fold cross-validation using gird-search to maximize root mean squared error (RMSE) on the training data. XGBoost is suitable as it effectively handles high-dimensional datasets by automatically capturing complex feature interactions and selecting the most important features. A total of 147 features (tax lines and company characteristics) are used to predict the level of non-compliance. The 10 most important features are shown in Figure 1B in the Appendix.

models SRC can decide to use the most efficient model for selecting companies for audit in 2025/2026 for the income year 2024.

34. **Selecting the top 2 percent highest risk scores according to the model is approximately half of the total number of audits SRC plans on conducting for the income year 2023.** With roughly 30,000 companies filing CIT returns in 2023, selecting the top 2 percent with the highest risk-scores according to the model is roughly equivalent to 600 corporations which is approximately half of the total number of audits SRC plans on conducting for the income year 2023 from July 1, 2024, to July 1, 2025. Given the typically high heterogeneity among corporations, it is suggested to allocate audits for the income year 2023 evenly between SRC's current strategy and the MLM, ensuring a 50/50 split. This increases the likelihood of drawing conclusions from statistically significant results. If too few audits (e.g., 10 percent or 20 percent for the income year 2023) are conducted using the MLM, there's a risk that differences between the two models will not be statistically significant. Consequently, it becomes impossible to determine which model is superior, resulting in a lack of useful information from the test.

35. **The IMF could assist SRC in further developing capacity for selecting cases for the income year 2023 using the MLM.** IMF can train and mentor SRC staff in selecting 50 percent of the cases for audits for the income year 2023 using the MLM.

36. **It is crucial that SRC updates the provided data with new data for the income year 2023 in the exact same format to facilitate the selection of cases for audit using the MLM**. Once corporations have filed their tax returns for the 2023 income year (deadline April 20, 2024), SRC should promptly update the data used in this mission. The data should be ready for updating the MLM before May 10, 2024.

37. **If the data is updated, the MLM can be used to select cases for audits for the income year 2023 before SRC publishes the list of corporations to be audited for income year 2023 on June 1, 2024**. If the data on the 2023 CIT returns is produced no later than May 10, 2024, SRC should be able to select cases and compile a list with taxpayer identification numbers (TINs) to be audited before May 20, 2024. This timeline ensures that SRC can prepare the list for publication on June 1, 2024.

C. Recommendations

Test cases selected from MLM

- Test the effectiveness of the MLM by selecting 50 percent of cases for income year 2023 using the traditional risk-based model and 50 percent of cases using the MLM.
- As a prerequisite for this, update the already delivered data with new data for income year 2023 in exactly the same way to select cases for audit for the income year 2023.
- Seek IMF capacity development (CD) to train staff and help select cases for income year 2023 using the developed MLM.
- Assess the results of the two risk selection methods and decide which model should be used for selecting audits for income year 2024.

- Examine the reasons for non-compliance more closely in high-risk sectors such as the mining sector to understand if it is due to deliberate evasion, lack of knowledge of the tax law or complicated legislation. This will give valuable information for a compliance improvement plan.

Data analytics

- Appoint a team of 2-3 people to engage in data analytics so that SRC can run the developed models itself.

- The appointed team should invest around 80 percent of their time on data analytics and learn the necessary programming tools to be able to carry out CIT Gap analysis and to maintain the MLM independently.

Annex I. Supplemental material

Tax gap estimates from non-random risk-based audits are prone to sample selection bias due to the selection process being influenced by the perceived risk of non-compliance. Put differently, the audited companies are not representative of the general population of companies since they are selected as being more prone to risk of tax non-compliance based on several indicators. Hence the tax gap estimate based purely on such operational risk-based audits does not reflect that of the general population. A common approach to account for this is through the Heckman 2-step estimator.[14] The Heckman 2-step estimator corrects for sample selection bias by estimating both the selection process and the level of non-compliance (i.e., the tax uncovered from audit) in the same model. Following Wooldridge (2010)[15] the Heckman 2-step estimator is given by an outcome equation and a selection equation:

$$Y_i = X_i\beta + u_i \qquad (1)$$

$$S_i = 1[Z_i\delta + v_i > 0] \qquad (2)$$

Here equation (1) is the outcome equation, where Y_i measures the tax uncovered from audit and X_i is company characteristics (i.e., lines from the CIT return, sector, number of employees etc.)[16]. Next, equation (2) is the selection equation, where S_i is an indicator of audit, with $S_i = 1$ denoting company i was audited and $S_i = 0$ denoting company i was not, while Z_i are factors that determine whether a company is audited or not. This includes 25 separate risk scores used by the SRC to target CIT audits.[17] Importantly, the value of Y_i is only observed if company i was selected for an audit ($S_i = 1$). Finally, u_i and v_i are independent of X_i and Z_i, with $v_i \sim N(0,1)$ and $E(u_i|v_i) = \gamma v_i$.[18]

What we are interested in estimating is $E[Y_i|X_i]$. However, since Y_i is observed only when $S_i = 1$, what we can estimate is $E[Y_i|X_i, S_i = 1]$. Using equation (1) and (2) this can be rewritten as:

$$E[Y_i|X_i, S_i = 1] = X_i\beta + E[u_i|v_i > -Z_i\delta] = X_i\beta + \gamma\lambda(Z_i\delta) \qquad (3)$$

Here $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density function (pdf) and cumulative distribution function (cdf) of a standard normal distribution, respectively. The form of $\lambda(\cdot)$ follows from the assumption that $v_i \sim N(0,1)$ and it is labeled the inverse Mills ratio. Equation (3) presents a way to

---

[14] James J. Heckman (1979). Sample Selection Bias as a Specification Error. *Econometrica*. vol. 47(1), pp. 153-161.

[15] Jefrey M. Wooldridge (2010). Econometric Analysis of Cross Section and Panel Data. The MIT Press

[16] A total of 25 variables are used. These variables correspond to the 25 variables with the highest variable importance in the machine learning model developed to predict audit adjustments. The top 25 were chosen because adding more variables led to limited increases in the explained variation of the audit adjustments and the risk of multicollinearity and loss of degrees of freedom.

[17] The 25 factors that determine each of the risk scores are confidential. Hence, the estimation of equation (2) was based on the assigned scores and not the underlying factors. Line item A?91 in section 1 of the corporation tax return "Total amount of profit tax deducted due to the privilege of deduction of profit tax" was not one of the 25 variables. This line item was included in a separate analysis after request from SRC but had not impact on the results due to the very limited number of corporations that filed an amount in line item A91. Similarly, after a request from SRC, an indicator defining whether a company is included in the group of resident taxpayers of the Republic of Armenia carrying out a program, and thereby subject to the privilege of lower tax rates, was included in a separate analysis. The inclusion also did not affect the results.

[18] We only require v_i to be normally distributed. It is sufficient to assume that the conditional expectation of u_i given v_i is linear, which does not require u_i to be normally distributed.

consistently estimate β. Following Heckman (1979) we can consistently estimate β and γ by regressing Y_i on X_i and $\lambda(Z_i\hat{\delta})$ using OLS, where $\hat{\delta}$ is obtained by estimating equation (2) using a probit model. Once an estimate of β has been obtained using the 2-step Heckman estimator, it can be used to construct an estimator of the unconditional expectation of non-compliance (not conditioning on $S_i = 1$), given by $E[\widehat{Y_i|X_i}] = X_i\hat{\beta}$.[19] This can be applied to obtain predicted values of non-compliance for all companies in the population, and thereby the overall CIT gap. Table A1 presents estimates of the selection and outcome models using all data on CIT returns and operational audits from 2020, 2021 and 2022. The selection model obtains an R^2 of 0.26.[20] This should be viewed in the context of the current audit strategy. When a company gets a high-risk score, it undergoes an audit. However, it's not just the income year when the company got the high score that gets audited; all previous non-audited CIT returns are also audited. Consequently, a company might have an income year audited even if it received a low score during that period, making risk-scores non-perfect predictors of audits at the TIN/income year level. Turning to the outcome model, it obtains an R^2 of 0.14. Ideally, we aim for this value to be as high as possible. However, due to the considerable diversity among companies and audit adjustments, reaching this goal is challenging. Interestingly, the coefficient on the inverse Mills-ratio is significant (IMR in Table A2), indicating the presence of sample selection bias.

Two important points need to be highlighted. First, when using the Heckman 2-step estimator to predict non-compliance instead of inferring causal relationships, the accuracy of predictions depends on how well the selection model and the outcome model explain the data. Second, it is best to avoid using the same variables in both models. Doing so makes the outcome model's identification rely on the non-linearity of the inverse Mills-ratio, which can cause unstable results due to high multicollinearity. To prevent this, the selection model should include at least one variable that determines whether a company gets audited but doesn't affect non-compliance levels (known as an exclusion restriction).[21] However, finding such a variable can be tricky if audits are solely based on estimated non-compliance. In this context, we use 25 risk scores to predict whether a company undergoes an audit. Table 3 in the main text shows that the relationship between the SRC's risk scores and average audit adjustments is non-monotone. This implies that some risk scores may have little to no connection to non-compliance levels, which satisfies the exclusion restriction.[22] However, if the SRC updates their targeting strategy and implements a MLM designed to identify large audit adjustments based on a large array of variables associated with the company, identifying exclusion restrictions may become more challenging.

[19] Standard errors are wrong when manually estimating the 2-step Heckman estimator. Correct standard errors can be obtained using bootstrap.

[20] This is McFadden's Pseudo R^2

[21] In other words, exclusion restriction means that there must be at least one variable appearing with a non-zero coefficient in the selection equation but not in the equation of interest.

[22] Regressing the audit adjustment on the risk scores reveals that 16 out of the 25 risk scores are not statistically significantly related to the tax uncovered from audit.

Table A2: Regression results

Selection Model			Outcome model		
Variables	**Coefficients**	**Std. Error**	**Variables**	**Coefficients**	**Std. Error**
Intercept	-1,115***	0,030	Intercept	6153,822**	3109,093
Risk Score 1	-0,005***	0,001	Tax line 85	-0,006	0,021
Risk Score 2	0,022***	0,001	Tax line 76	0,001	0,005
Risk Score 3	-0,006***	0,001	Tax line 50.5	-0,066	0,048
Risk Score 4	-0,001*	0,000	Turnover pr. emp.	0,003	0,002
Risk Score 5	-0,003***	0,000	Tax line 84	0,010	0,022
Risk Score 6	-0,006***	0,001	Number of emp.	11,705**	4,015
Risk Score 7	-0,001**	0,001	Tax line 40	-0,001	0,001
Risk Score 8	0,007***	0,001	Tax line 83	-0,005	0,006
Risk Score 9	-0,007***	0,001	Tax line 19	0,183	0,127
Risk Score 10	0,005***	0,002	Tax line 7	0,001	0,001
Risk Score 11	-0,001***	0,000	Tax line 62	-0,005	0,006
Risk Score 12	0,002***	0,000	Tax line 50	-0,001	0,001
Risk Score 13	0,002	0,002	Tax line 50.4	-0,150**	0,061
Risk Score 14	0,000	0,000	Total compensation	-0,031*	0,016
Risk Score 15	-0,004***	0,001	Tax line 11	0,000	0,001
Risk Score 16	-0,014***	0,001	Tax line 59	0,005	0,004
Risk Score 17	-0,002	0,002	Tax line 59	0,006	0,012
Risk Score 18	-0,010***	0,001	Tax line 51	0,000	0,002
Risk Score 19	0,005***	0,001	Tax line 58	-0,248	0,283
Risk Score 20	0,005***	0,001	Tax line 16	0,007	0,007
Risk Score 21	0,009***	0,002	Tax line 41	0,005	0,021
Risk Score 22	-0,010***	0,001	Tax line 43	0,001	0,001
Risk Score 23	0,041***	0,003	Tax line 9	0,000	0,001
Risk Score 24	-0,026	0,058	Tax line 45	0,000	0,001
Risk Score 25	-0,021***	0,002	IMR	-1032,997*	578,905
Includes Sector dummies	No			Yes	
Number of observations	71,374			4,324	
R^2	0.26			0.14	

Source: IMF calculations based on data from SRC. Note: [1)] In 2020-prices. 0.1 pct. of audit results were trimmed in top and bottom for each year. Re-audits are discarded. Only corporations with a reported CIT return. The outcome model also includes sector dummies. Standard errors are computed using bootstrap, with resampling done at the TIN-level. For the selection model R^2 corresponds to McFadden's Pseudo R^2. *, **, *** denotes p < 0.01, p < 0.05, p<0.10.

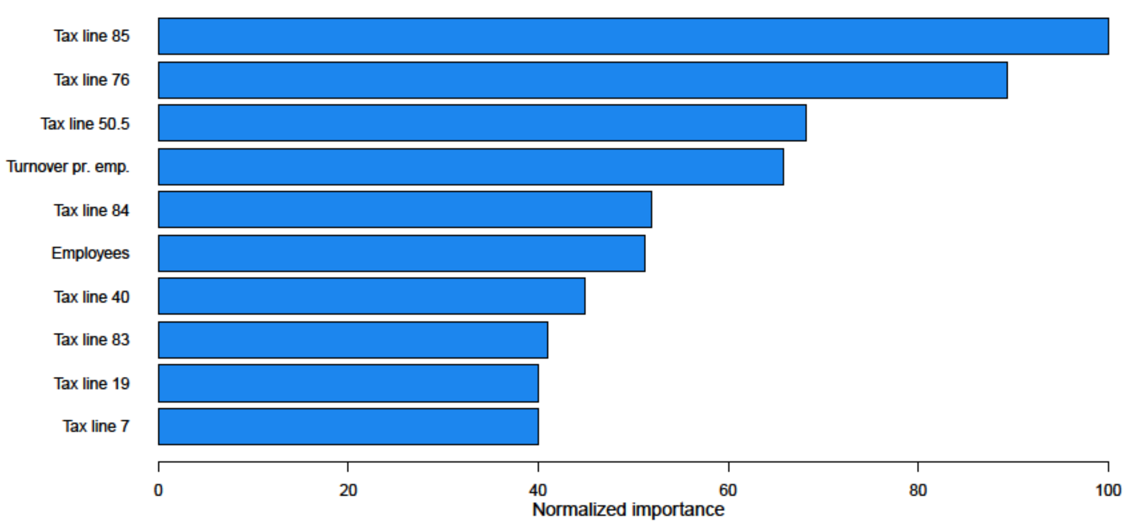
A. Supplementary results

Table 1B. CIT Gap estimates based on operational audits (excluding the largest corporations)

	Panel A: Using all audit adjustments			
	Overall	**2020**	**2021**	**2022**
CIT gap	64.9 billion AMD[1)]	57.1 billion AMD[1)]	62.5 billion AMD[1)]	75.0 billion AMD[1)]
Percent of potential CIT	34.2	36.9	33.3	33.1
Percent of GDP	1.0	0.9	1.0	1.0
	Panel B: Excluding audit adjustments with no immediate effect on revenue			
	Overall	**2020**	**2021**	**2022**
CIT gap	47.5 billion AMD[1)]	41.4 billion AMD[1)]	45.8 billion AMD[1)]	55.4 billion AMD[1)]
Percent of potential CIT	27.9	30.0	27.0	26.9
Percent of GDP[2)]	0.7	0.7	0.7	0.8

Source: IMF calculations based on data from SRC. Note: 1) In 2020-prices. 0.1 pct. of audit results were trimmed in top and bottom for each year. Re-audits are discarded. The top 0.1 percent of the largest companies, as measured by turnover, were trimmed each year. Only corporations with a reported CIT return. Potential CIT liability is defined as self-reported CIT plus the estimated CIT gap. Panel A presents estimates of the CIT gap using all audit adjustments. Panel B excludes audit adjustments with no immediate effect on revenue. 2) GDP is also in 2020 prices in the calculation of the CIT gap in percent of GDP.

Figure 1B. Variable importance of the MLM



Source: IMF calculations based on data from SRC. Note: Variable importance is measured by "Gain," which measures the reduction in loss achieved by splitting on a particular variable. The variable importance has been normalized with respect to the most important variable.

B. Materials Left with Armenia State Revenue Committee

- Code in R with thorough explanations for measuring the CIT Gap based on operational audits.
- Code in R with thorough explanations for the MLM.
- Code in R that collects and cleans data.
- Code in R that calculates summary statistics based on operational audits.
- Code in R that calculates bootstrap standard errors for the Heckman model.
- Powerpoint presentation explaining the method used to correct for “Sample Selection Bias” used at opening meeting.
- Powerpoint presentation that shows how to implement Heckman method step-by-step.
- Powerpoint presentation with main results from the mission presented at exit meeting.
- List with 1,000 (anonymized) TIN’s with highest risk scores according to the MLM for the income year 2022.
- List with 1,000 (anonymized) TIN’s with highest risk scores according to SRC’s current risk model for the income year 2022.
- List with variable importance measures according to MLM.