European Commission

VAT COMPLIANCE GAP DUE TO MISSING TRADER INTRACOMMUNITY (MTIC) FRAUD

FINAL REPORT - PHASE I

CASE – Center for Social and Economic Research (Project Leader)
WIFO – Austrian Institute of Economic Research (Consortium Leader)
PwC (Contributor)

Taxation and Customs Union

EUROPEAN COMMISSION

Directorate-General for Taxation and Customs Union
Directorate C – Indirect taxation and tax administration
Unit C.5 – Economic analysis and taxation of exempted sectors

E-mail: TAXUD-UNIT-C5@ec.europa.eu@ec.europa.eu

European Commission
B-1049 Brussels

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*): 00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you)

More information on the European Union is available on the Internet (http://www.europa.eu).

Luxembourg: Publications Office of the European Union, 2024.

PDF	ISBN 978-92-68-13299-9	doi 10.2778/25011	KP-05-24-169-EN-N

This document should be referred to as
European Commission, CASE, Poniatowski, G., Śmietanka, A., Skowronek, A., Sojka, A., Orzechowski, K., *VAT compliance gap due to Missing Trader IntraCommunity (MTIC) Fraud – Final Report Phase I*, Publications Office of the European Union, Luxembourg, 2024

Acknowledgments

This report was written by a team of experts from CASE (Center for Social and Economic Research, Warsaw) directed by Grzegorz Poniatowski, and composed of Adam Śmietanka, Agnieszka Skowronek and Aleksandra Sojka. Research assistance was provided by Krzysztof Orzechowski. The Project was coordinated by Simon Loretz (Austrian Institute of Economic Research, Vienna) and Magdalena Wiśniewska (CASE, Warsaw). PricewaterhouseCoopers LLP (UK) provided research and analysis regarding MTIC fraud and government approaches to reduce MTIC in preparation of the report.

We acknowledge valuable comments from reviewers, Silvia Fedeli and Michael Udell, and thorough proofreading by Kristen Hartwell. We also acknowledge valuable support of officials of tax and statistical offices of the Member States, who shared information, comments, and suggestions. All responsibility for the estimates and the interpretation in this report remains with the authors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of acronyms and abbreviations

AU-ROC	Area Under the Receiver Operating Characteristic Curve
B2B	Business-to-Business
B2C	Business-to-Consumer
BEC	Broad Economic Categories
CASE	Center for Social and Economic Research (Warsaw)
CIF	Cost of Insurance and Freight
CIT	Corporate Income Tax
CN	Combined Nomenclature
CPA	Statistical Classification of Products by Activity
EC	European Commission
EPPO	European Public Prosecutor's Office
EU	European Union
EY	Ernst & Young
FOB	Free on Board
FTE	Full Time Equivalent
GATS	General Agreement on Trade in Services
GDP	Gross Domestic Product
HMRC	His Majesty's Revenue & Customs
HS	Harmonised System
ICA	Intra-Community Acquisition
ICS	Intra-Community Supply
IMEI	International Mobile Equipment Identity
IOSS	Import One-Stop Shop
KYC	Know Your Customer
LROC	Localization Receiver Operating Characteristic
MIMIC	Multiple Indicators, Multiple Causes Measurement
MTEC	Missing Trader Extra-Community
MTIC	Missing Trader Intra-Community
MS	Member State
NACE	Statistical Classification of Economic Activities in the European Community
OECD	Organisation for Economic Co-operation and Development
OSS	One-Stop-Shop
PIT	Personal Income Tax
PwC	PricewaterhouseCoopers LLP (UK)
RCM	Reverse Charge Mechanism
ROC	Receiver Operating Mechanism
SAF-T	Standard Audit File for Tax
SCM	Standard Cost Model
SEM	Structural Equation Modelling

STIC	Standard International Trade Classification
SUT	Supply and Use Tables
TARC	Tax Administration Research Centre
TAXUD	Directorate-General for Taxation and Customs Union
TNA	Transaction Network Analysis
UNCTAD	United Nations Conference on Trade and Development
VAT	Value Added Tax
VDA	Virtual Digital Assets
ViDA	VAT in the Digital Age
VIES	Vat Information Exchange System
VoIP	Voice over Internet Protocol
WTO	World Trade Organisation
XML	Extensible Markup Language

Introduction

The European Commission has been regularly monitoring VAT collection efficiency in the EU since the 2013 *Study to quantify and analyse the VAT Gap in the EU-27*. From this moment, reports presenting updated analyses of the VAT compliance and policy gaps have been published annually. Overall, following the publication of the 2023 report, fully fledged VAT gap estimates now cover the period of 2000-2021.

The *VAT gap in the EU* studies have provided policy makers with essential knowledge about the scale of revenue forgone due to non-compliance and the design of VAT rules. Using a standardised methodology and data sources, the studies allowed for comparisons across time and against other Member States. The estimates have also served as a useful tool to help understand the nature of forgone revenue and provide insights on the strategies that improve the efficiency of VAT collection. To achieve that, the studies have taken advantage of the availability of consistent estimates of time and cross-country differentiation, as well as statistical and econometric methods used for panel data analysis.

The reports published by the Commission have continuously refined the methodological approach in order to increase the precision of the estimated theoretical tax liabilities, which are the reference point for the VAT gaps calculation. The well-established methodological approach used by the study, the so-called top-down consumption-side method, could be characterised by a number of advantages, including simplicity, cross-country comparability and low dependence on non-publicly available data sources. Yet, it also poses some limitations. Importantly, it does not allow one to break the VAT compliance gap by sectors of economic activity and types of irregularities. As the forms of non-compliance are numerous, ranging from the legal exploitation of loopholes in tax systems to evasion or organised large-scale tax fraud, knowledge of the relevance of these components could be useful for improving policy decisions and the functioning of tax administrations in the EU.

The form of non-compliance in VAT in the EU that deserves special attention is missing trader fraud. Missing trader fraud refers to a scheme in which a fraudulent trader supplies goods and services to other businesses, collects the tax due on the supply from their customers, and disappears without ever remitting it to the tax authorities. The European Commission defines a missing trader as "a trader registered as a taxable person for VAT purposes who, potentially with a fraudulent intent, acquires or purports to acquire goods or services without payment of VAT and supplies goods and services with VAT but does not remit the VAT due to the appropriate national authority".[1] This type of fraud can target different taxes and types of transactions, since there are multiple circumstances under which not paying tax and disappearing could prove very profitable.

MTIC fraud is a specific type of VAT fraud which involves taking advantage of the fact that intra-community movement of goods and services is VAT-free, resulting in the fraud being more profitable.[2] In recent years, MTIC fraud was likely one of the main sources of VAT non-compliance. According to Europol (2013), MTIC fraud accounted for over EUR 100 billion VAT loss in the EU-28 annually.[3] Although more recent studies quote more conservative estimates,[4] there is a broad consensus in the

[1] FISCALIS (2018).

[2] Ibid.

[3] Source: https:/www.europol.europa.eu/sites/default/files/documents/socta2013.pdf.

[4] See e.g., Braml and Felbermayr (2021).

literature that revenue forgone due to MTIC fraud continues to make up a substantial share of the entire VAT compliance gap, i.e. VAT revenue lost due to not meeting legal obligations.[5]

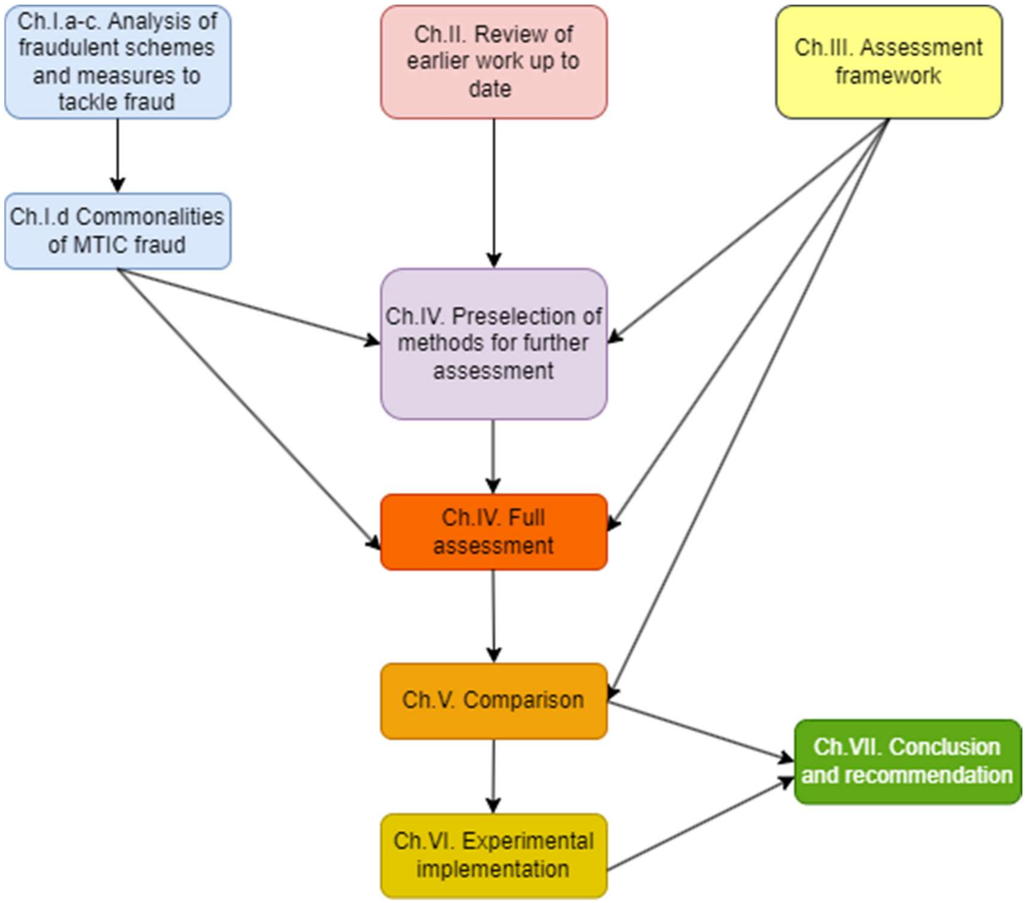
This report has been prepared for the European Commission, DG TAXUD, with a focus on the *VAT compliance gap due to Missing Trader Intra-Community (MTIC) fraud* (henceforth referred to as the "MTIC gap study") under Framework Service Contract No. TAXUD/2020/CC/159. It presents the final results of Phase I of the study, which aims to provide the best-suited methodological and operational framework that could be employed for recurrent yearly estimation of VAT revenue foregone due to MTIC fraud. The estimates using this methodology will complement the estimates of the overall VAT compliance gap and will be published in a dedicated report after the completion of Phase II.

The report consists of seven chapters, structured in accordance with the reasoning flow presented in Figure 1. The first chapter analyses the logic and course of action of the MTIC fraud and its subtypes, which is a prerequisite for defining the concepts and the scope of work. The second chapter summarizes work on estimating the scale of fraud conducted up to date. It reviews publicly available reports, research papers and articles with the objective of mapping existing methodological approaches and describing their characteristics. Building on that, Chapter II analyses the experiences of Member States' administrations with operationalizing the methods enumerated in the preceding chapter and not reported in the public domain. The third chapter introduces the framework employed for the (i) pre-selection of methodological approaches, (ii) their full assessment and (iii) comparison. Chapter IV discusses the first step in the selection process – the pre-assessment and preselection of methods and their grouping into broader methodological scenarios. This step is taken using the knowledge gathered from secondary sources (Chapter II), complemented by the data availability and methodological gaps analysis presented in the same chapter. Chapter V assesses the criteria envisaged by the assessment framework for each of the scenarios, and Chapter VI provides a comparison of the scores of all scenarios. Chapter VII concerns the experimental implementation of two of the proposed methodological scenarios. The final chapter summarizes the findings and gives provides recommendations.

---

[5] See: https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/690462/IPOL_BRI(2021)690462_EN.pdf.

Figure 1: Structure and reasoning flow in the report



Source: own elaboration.

This report comes with seven appendices. Appendix A contains the glossary. Appendix B, provides supplementary information for the mapping of MTIC fraud pathways.[6] Appendix C discusses the assumptions underlying the assessment framework. Appendix D describes the analytical methods referenced in the main body of the report. Appendix E summarizes the responses to the questionnaire for tax and statistical authorities. Appendix F presents the questionnaire itself. Appendix G enumerates and describes revealed cases of MTIC fraud, which are useful for understanding the characteristics of fraud. Appendix H contains tables with comments to the Inception Report and Draft Final Report from the external reviewers and the authors' responses to the points raised.

[6] Appendix B was excluded from the published version of the report.

I. Mapping of MTIC fraud pathways

MTIC fraud can take different forms and vary in magnitude, ranging from simple acquisition fraud to complex networks spanning multiple Member States and involving actors not necessarily aware of their participation in a fraudulent chain of transactions. This chapter analyses different archetypes of MTIC fraud and methods used to obfuscate detection of fraud by administrations. The analysis presented herein is a prerequisite for defining MTIC fraud and, in consequence, the scope of this study, summarized in *I.d. Commonalities across different MTIC schemes.* As differences across different subtypes of schemes are related to how the fraud is recorded in various datasets, the information presented in this chapter is also crucial for the assessment of alternative methodological scenarios, their completeness and ability to break down the estimates by basic types of schemes.

There are four common archetypes of MTIC fraud often distinguished by the literature. These are: *simple acquisition*, *carousel*, *contra-trader* and *cross-invoicer fraud*.[7] The distinction of multiple archetypes is clearly related to the complexities that fraudsters add to the MTIC schemes to avoid detection. Yet, as explained in more detail in the following sections, through the lens of the objectives of this study it is more practical to distinguish two basic archetypes – *simple acquisition fraud* and *carousel fraud* – and analyse how different means used to complicate the detection of fraud affect the ability to detect it in various datasets. Despite their diversity, the schemes presented in this chapter always involve some combination of the following actors[8]:

- **Missing trader:** As set out earlier, a missing trader is "a trader registered as a taxable person for VAT purposes who, potentially with a fraudulent intent acquires or purports to acquire goods or services without payment of VAT and supplies goods and services with VAT but does not remit the VAT due to the appropriate national authority" (for example, Company B in Figure 2 and Figure 3).[9] In a fraudulent transaction chain, the missing trader is typically the company carrying out the intra-Community acquisition, benefitting from the EU rule that cross-border movement of supplies is VAT free. In most instances the individuals running the "Missing Trader" entity are aware of the scheme. However, in some instances an innocent company may have their VAT registration number hacked by criminals.
- **Broker company:** A broker company typically sits at the end of the transaction chain in the Member State in which the missing trader is situated. It purchases the goods and services from either the missing trader or another business in the transaction chain (a buffer company, see below), and then sells them to a business in another EU Member State (for example Company C in Figure 3). The broker company is typically a willing participant of the fraudulent transaction chain.
- **Buffer company:** A buffer company is a normal trader, placed in the transaction chain. They are typically placed between the missing trader and the broker company to make it harder to detect the scheme. More than one buffer company can be added to the scheme

[7] See e.g., FISCALIS (2018).

[8] Definitions sourced from FISCALIS (2018).

[9] Commission Regulation (EC) No 1925/2004 of 29 October 2004 laying down detailed rules for implementing certain provisions of Council Regulation (EC) No 1798/2003 concerning administrative cooperation in the field of value-added tax, Article 2, https://eur-lex.europa.eu/legalcontent/EN/ALL/?uri=celex%3A32004R1925.

to make it more complex (for example Company C in Figure 4). The buffer company does not necessarily need to be aware of its participation in a fraudulent chain of transactions.

- **Conduit company:** A businesses based in one Member State and selling to businesses (or final consumer) in other Member States (intracommunity acquisitions followed by intracommunity supplies). A company is qualitied as a conduit if it was involved in a fraudulent transaction (for example Company A in Figure 2 and Figure 3). The conduit company does not necessarily need to be aware of its participation in a fraudulent chain of transactions.

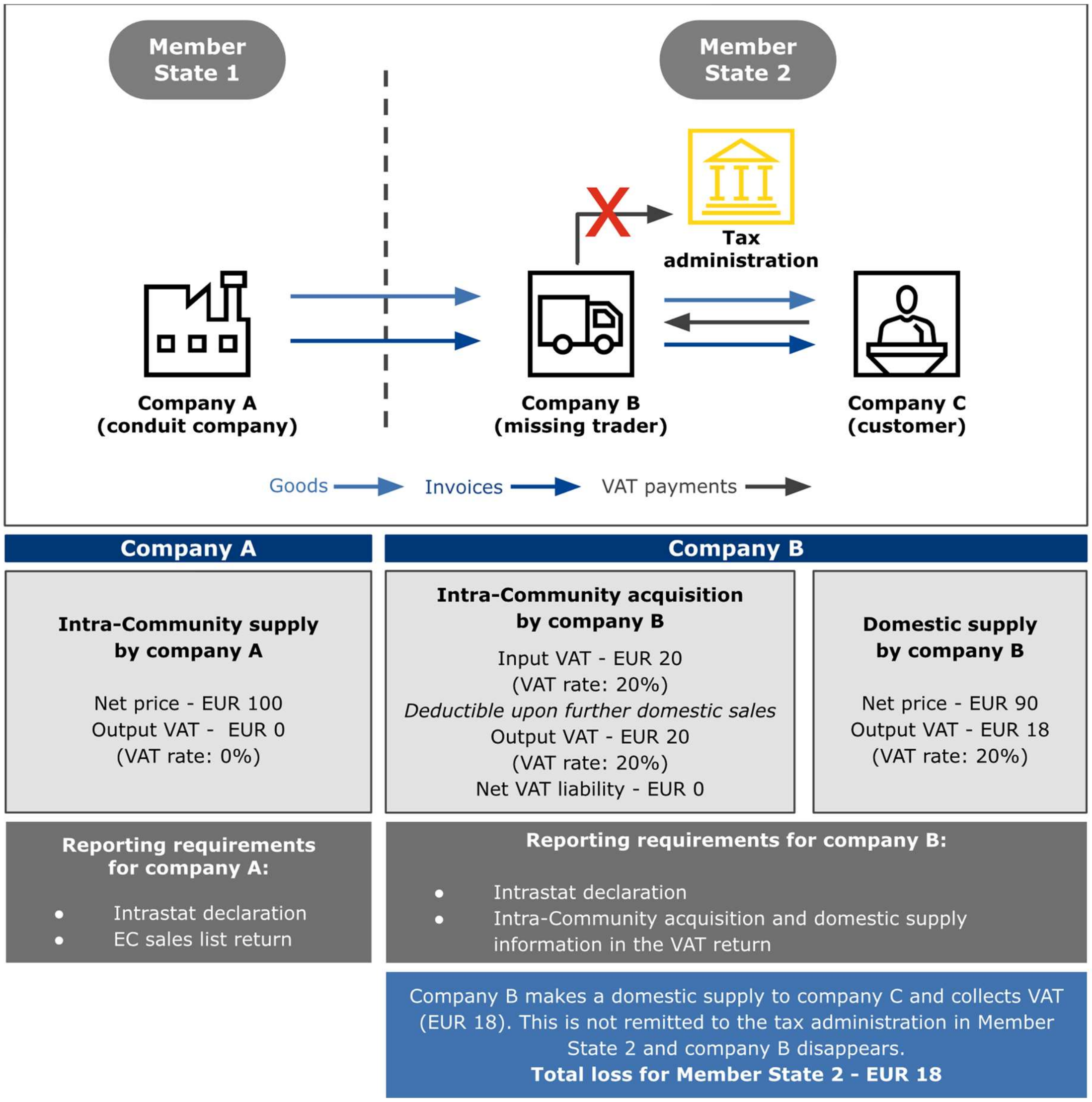
I.a. Archetypes of MTIC fraud

Simple acquisition fraud

Being the simplest MTIC scheme, acquisition fraud relies on the fact that, in order to avoid double taxation, VAT is not charged on cross-border transactions between two or more EU Member States (MS). In its simplest form, MTIC fraud involves a fraudulent company (*Company B* in Figure 2) which purchases goods and services from a company (*Company A*) in another EU Member State (Member State 1) under the so-called *reverse charge*. This means that, contrary to standard VAT treatment in domestic transactions, VAT is remitted by the acquirer rather than by the supplier.[10] *Company B* then charges VAT on the subsequent sale to *Company C*, and disappears without remitting the VAT collected to the tax administrations in Member State 2.

[10] As discussed in more detail in Section I.c, the reverse charge can also be applied to selected domestic transactions. This mechanism is introduced by administrations to prevent missing trader fraud, in cases where purchasers are more likely to comply with the VAT payment obligation.

Figure 2: Simple MTIC acquisition fraud



Source: own elaboration based on FISCALIS (2018).

In this scenario, the company making the Intra-Community Supply (ICS, *Company* A) is likely to submit an Intrastat declaration[11] and an EC sales list return[12] outlining the value of the supply (EUR 100 in the simple numerical example in Figure 2) and the VAT registration details of the acquirer (*Company B*). With the introduction of the EU VAT quick fixes in 2020, suppliers such as *Company A* are also required to obtain and validate the VAT number of the acquirer (in this case *Company B*) and quote it on their invoice, along with submitting the EC sales list returns in order to make a zero-rated intra-community supply.[13] The acquirer (*Company B*), who will be acting as a missing trader, is required to

[11] Note: Only businesses that meet certain thresholds for intra-community supplies and acquisition are required to submit Intrastat declarations. Intrastat declarations are also only required for goods. Further information on Intrastat declarations can be found in Section IV.a.

[12] With the introduction of the EU VAT quick fixes in 2020, suppliers are required to capture the intra-community supply in their EC sales listing in order for it qualify for the application of zero VAT rate. Further information on EC sales list return can be found in Section IV.a.

[13] See: https://www.pwc.nl/en/insights-and-publications/tax-news; https://www.avalara.com/vatlive/en/vat-news.

file an Intrastat declaration for the Intra-Community Acquisition (ICA) and a VAT return.[14] The VAT return should include the Intra-Community Acquisition and subsequent domestic supply. The VAT owed as part of the Intra-Community Acquisition is deductible and will therefore be recorded as both input (EUR 20) and output VAT (EUR 20) by the acquirer/missing trader (*Company B*) in their VAT return. The acquirer/missing trader (Company B) then makes a subsequent sale to *Company C*, typically at a lower price (assumed to be EUR 90 in the simple numerical example set out in Figure 2) in order to make the goods/service more commercially attractive than those being offered by other traders in the market. It is possible for the acquirer/missing trader (*Company B*) to sell below the initial purchase price (EUR 100 in this example) because the illegal margin that is made from collecting VAT on the domestic supply and failing to remit (EUR 18 in this example) is greater than the loss on the underlying supply (EUR 10 in this example). The VAT that the acquirer/missing trader (*Company B*) collects from the domestic supply (EUR 18) should be shown as output VAT in their VAT return and be remitted to the tax administration.[15] An assumption could be made that, in many instances, missing traders are not likely to file a VAT return and Intrastat declaration, since they do not plan on remitting the tax to the tax administration. Yet, the behaviour of fraudsters depends on whether they deem the discrepancy in mirror registers or the reporting of high-value Intra-Community Acquisition to be more likely to alert the administration and trigger an audit. This expectation is likely driven by the track record of the administration and the evolution of the means of exchanging information and tracking fraud.

Following the domestic transaction, *Company B* will disappear with the VAT. Member State 2 incurs a VAT loss equivalent to the tax that the acquirer/missing trader (*Company B*) collects from the domestic supply transaction to *Company C* (EUR 18) and does not remit to the tax administration. The supplier's (*Company A*) Intrastat declarations and EC sale returns will contain information on the missing trader (*Company B*) and the value of the goods and services sold to them. This can potentially be used as a basis to estimate the VAT loss.

Carousel fraud

Carousel fraud (see Figure 3) is an extension of simply acquisition fraud, one of the more basic forms of MTIC fraud. Simple carousel frauds have a long history and remain a problem today. For instance, in 2021 a raid uncovered a scheme involving luxury cars and spanning several Member States, estimated

[14] Further information on VAT returns can be found in Chapter IV.a.

[15] In the simple numerical example set out in Figure 1, VAT in Member State 2 is charged at 20%.

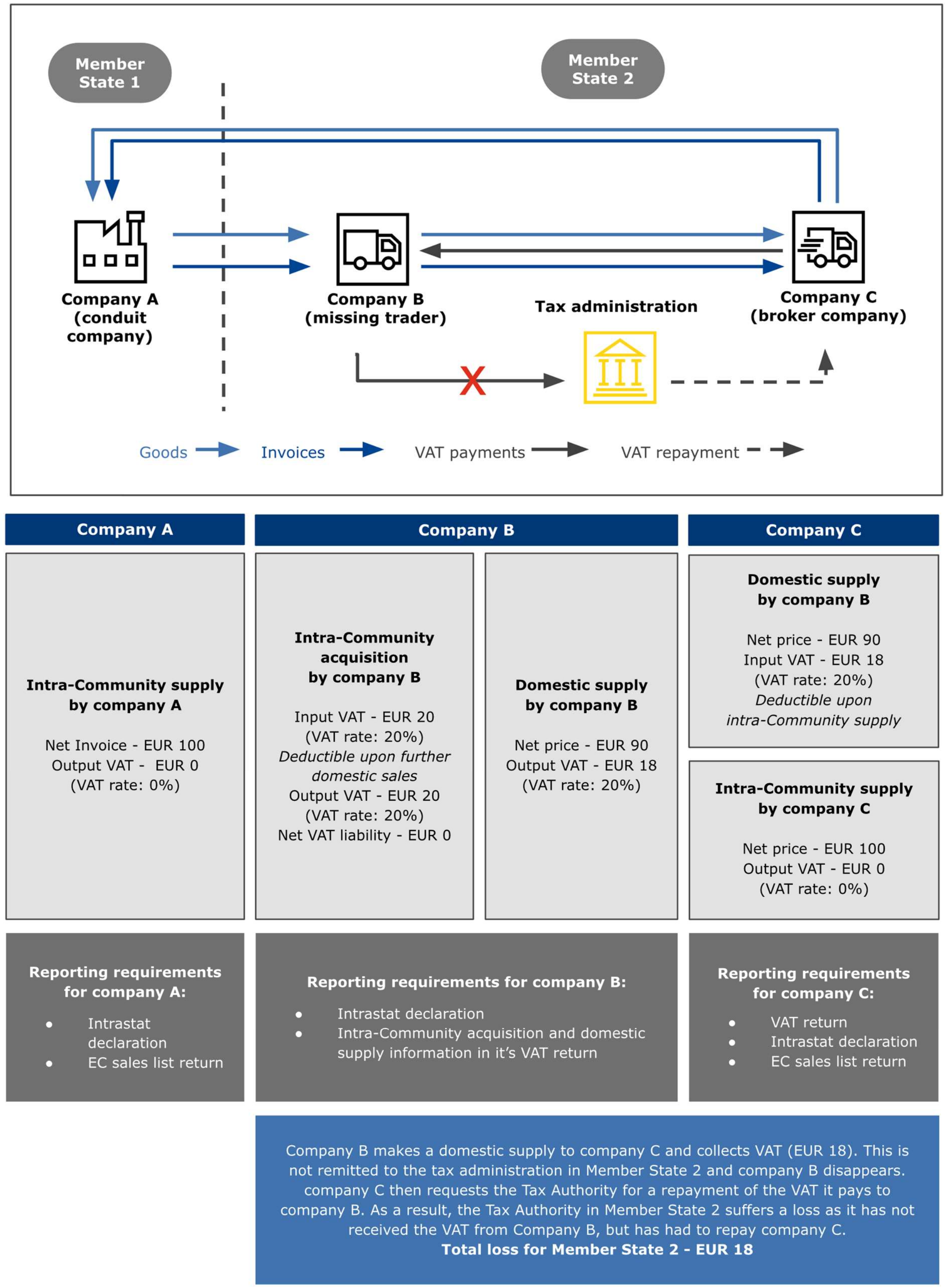
to have cost over EUR 13 million in lost revenue and led to the conviction of 10 individuals for tax evasion and drug trafficking.[16]

Similar to acquisition fraud, the missing trader (*Company B* in Figure 3) purchases goods and services from a conduit company (i.e., *Company A*) in another EU Member State (at EUR 100 in the below example), and subsequently carries out a domestic supply transaction to the broker company (*Company C*, at EUR 90) and charges VAT (EUR 18).[17] It then disappears without remitting the VAT to the tax authorities. The broker company (*Company C*) carries out an Intra-Community Supply back to the initial supplier, the conduit company (*Company A*) in Member State 1 (at EUR 100 in the below Figure), rather than supplying the goods and services further to an end consumer in Member State 2. As a result, *Company C* is able to reclaim from the tax administration in Member State 2 the VAT it has paid on the purchase of goods and services (i.e., input VAT). Once the supplies are resold to the conduit company (*Company A*), the fraud can be repeated multiple times between the same companies, like a carousel, which distinguishes this scheme from simple acquisition fraud.

[16] VAT Update (2022). First conviction in VAT carousel fraud involving luxury cars.

[17] In the simple numerical example set out in Figure 2, VAT is charged at 20%.

Figure 3: MTIC carousel scheme



Source: own elaboration based on FISCALIS (2018).

Member State 2 incurs a VAT loss of EUR 18 on account of the missing trader (*Company B*) disappearing without remitting the VAT it collected from its domestic supply transaction. The conduit company (*Company A*) is likely to act in a legitimate manner and will therefore submit an Intrastat declaration[18] and an EC sales list return.[19] These documents will include details on the missing trader (*Company B*) (i.e., the name of the company and its VAT registration number) and the value of the goods and services supplied (EUR 100 in our example). It is worth noting that there are some differences in the reporting requirements for Intrastat declarations and EC sales list returns, which are further outlined in Appendix A and Chapter IV. The conduit company (*Company A*) will also be required to validate the VAT registration number of the missing trader (*Company B*) and include this in its invoice.[20]

The missing trader (*Company B*) is required to make an Intrastat declaration (for its Intra-Community Acquisition of EUR 100) and a VAT return,[21] including information on the Intra-Community Acquisition (including the EUR 20 of VAT under the reverse charge for which a deduction of input VAT can be claimed) and domestic supply (valued at EUR 90 in this example with output VAT of EUR 18).[22] However, they are less likely to submit this return, as there is no intention of remitting the VAT collected. Instead, the missing trader (*Company B*) will disappear with the VAT collected from the domestic supply (EUR 18).

The broker company (*Company C*) is also required to submit a VAT return. This will include the VAT it has paid to the missing trader (*Company B*) as input VAT (in this case, EUR 18), which it will then be able to deduct, as it is making an Intra-Community Supply to the conduit company (*Company A*). The broker company (*Company C*) will also submit Intrastat declarations and EC sales list returns, which will capture information on the movement of the goods and services back to the conduit company (*Company A*), based on a supply of EUR 100 and no VAT due (assuming the reverse charge is applicable to most intra-Community transactions). In addition, the broker company (*Company C*) will validate the VAT registration number of the conduit company (*Company A*) and quote it in the invoice. There is an interest in correct filing of records on the part of the broker company (*Company C*) in order to reclaim the input VAT and to give the impression of a legitimate trader.

Overall, these transactions and reporting incentives could "inflate" Intra-Community Acquisition and Intra-Community Supply transaction data of the broker and conduit companies, but not necessarily the Intra-Community Acquisition associated with the missing trader.

I.b. MTIC fraud in practice

While there are only two basic mechanisms of MTIC fraud, in practice the parties involved in MTIC fraud use a range of other techniques and schemes to maximise revenue and hinder their detection, as set out below.

Additional intermediaries. As shown in the MTIC mechanism diagrams above, the schemes may include different parties, such as a buffer company, broker company, and conduit company. These companies are intrinsic to certain forms of fraud. Adding more than one buffer, broker, and/or conduit company to a scheme makes it more complex and, in turn, harder to detect. Notably, some of the

[18] Note: Only businesses that meet certain thresholds for intra-community supplies and acquisition are required to submit Intrastat declarations. Intrastat declarations are also only required for goods. Further information on Intrastat declarations can be found in Section IV.a.

[19] With the introduction of the EU VAT quick fixes in 2020, suppliers are required to capture the intra-community supply in their EC sales listing in order for it qualify for the application of zero VAT rate. Further information on EC sales list return can be found in Section IV.a.

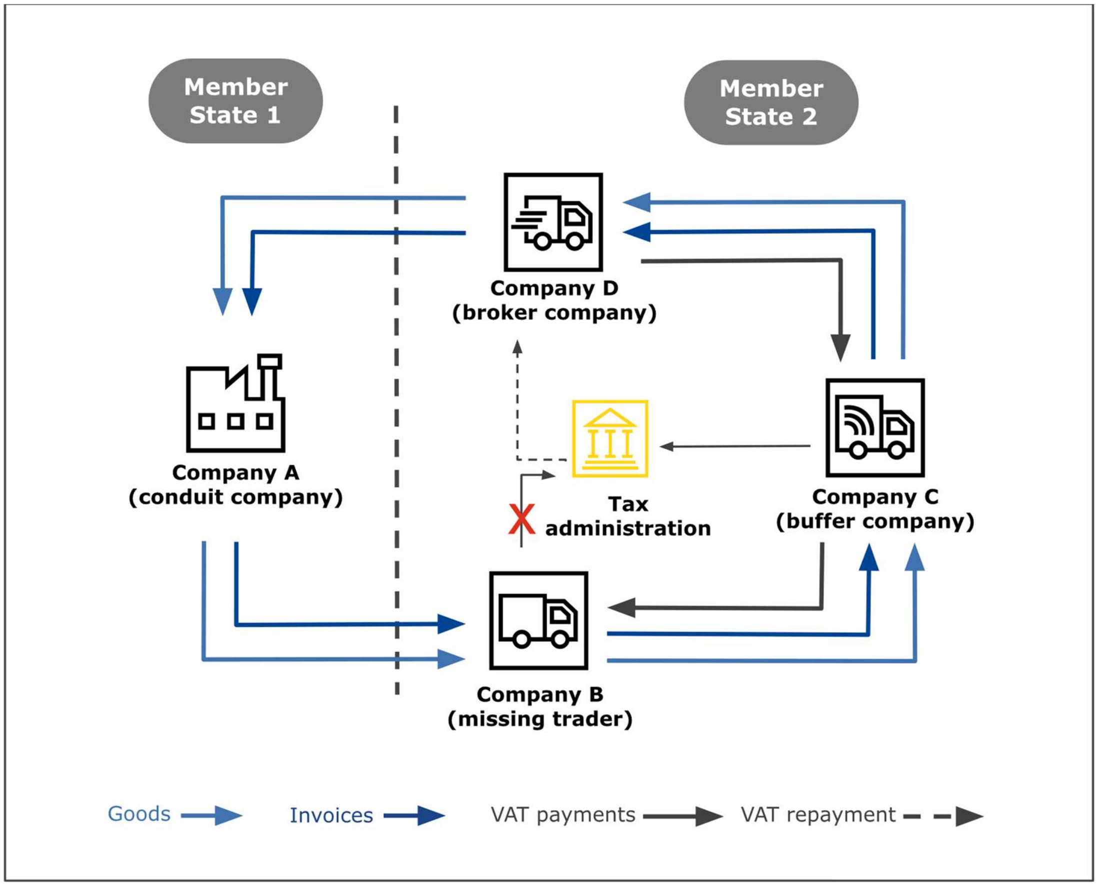
[20] *See:* https://www.pwc.nl/en/insights-and-publications/tax-news; https://www.avalara.com/vatlive/en/vat-news.

[21] Further information on VAT returns can be found in Section IV.a.

[22] The reverse charge in intra-Community transactions should not be confused with the domestic reverse charge mechanism (see Section I.c).

additional intermediaries involved in these schemes may not always be aware of being part of a wider fraud network. This might further inhibit detection. In discussions with tax practitioners, it was noted that long chains, consisting of perhaps five or six intermediaries before reaching the fraudulent company, were common. Figure 4 therefore shows a simple example of a carousel with an additional buffer company, but in practice there can be many additional intermediaries.

Figure 4: Carousel fraud with a buffer company



Company D

Domestic supply by company C
Net price - EUR 100
Input VAT - EUR 20 (VAT rate: 20%)
Deductible upon intra-Community supply

Intra-Community supply by company D
Net price - EUR 100
Output VAT - EUR 0 (VAT rate: 0%)

Reporting requirements for company D:

- Intrastat declaration
- EC sales list return
- VAT return

Company A

Intra-Community supply by company A
Net Invoice - EUR 100
Output VAT - EUR 0 (VAT rate: 0%)

Reporting requirements for company A:

- Intrastat declaration
- EC sales list return

Company C

Domestic supply by company C
Net price - EUR 100
Output VAT - EUR 20 (VAT rate: 20%)

Domestic supply by company B
Net price - EUR 90
Input VAT - EUR 18 (VAT rate: 20%)

Reporting requirements for company C:

- VAT return

Company B

Intra-Community acquisition by company B
Input VAT - EUR 20 (VAT rate: 20%)
Deductible upon further domestic sales
Output VAT - EUR 20 (VAT rate: 20%)
Net VAT liability - EUR 0

Domestic supply by company B
Net price - EUR 90
Output VAT - EUR 18 (VAT rate: 20%)

Reporting requirements for company B:

- Intrastat declaration
- Intra-Community acquisition and domestic supply information in it's VAT return

Company B makes a domestic supply to company C and collects VAT (EUR 18). This is not remitted to the tax administration in Member State 2 and company B disappears. Company D then requests the Tax Authority for a repayment of the VAT it pays to Company C. As a result, the Tax Authority in Member State 2 suffers a loss as it has not received the VAT from company B, but has had to repay company D.
Total loss for Member State 2 - EUR 18

Source: own elaboration based on FISCALIS (2018).

Additional Member States involved in a transaction. In addition to further buffer companies, there can be more than two or three Member States involved in the main transaction, as shown throughout this chapter (see e.g., Figure 4). Supplies could pass through multiple Member States in a single carousel scheme, as in the case of a recently discovered scheme involving SD cards.[23] MTIC schemes can also involve triangular trade, which is legal in principle, in order to make it harder to detect the fraud.[24] In a triangular trade setup, a supplier (*Company A*) in one Member State sells the goods and services to a business (*Company B*) in another Member State. This business (*Company B*) then goes on to sell the goods and services to another business (*Company C*) in a third Member State. However, the goods and services are delivered directly from *Company A* to *Company C.*[25] Introducing triangular

[23] See: https://www.europol.europa.eu/media-press/newsroom/news/vat-fraud-clampdown-international-scam-memory-cards-uncovered-in-netherlands.
[24] See: FISCALIS (2018).
[25] See: https://www.revenue.ie/en/vat/goods-and-services-to-and-from-abroad/intracommunity-supplies/what-is-triangulation

trade in an MTIC scheme adds an additional layer of disjointed invoicing and creates a "good chain", thus making the fraud harder to detect.[26] Through doing so, MTIC crime groups exploit the deficiencies in information sharing between national authorities – particularly when different types of institutions are concerned (e.g. tax administrations and law enforcement agencies) – and with the EU institutions (e.g. OLAF, EPPO).

Extra-community transactions. The example schemes given above embrace only intra-Community transactions, but fraud chains may also have an extra-EU element. Missing trader extra-Community (MTEC) fraud can occur between countries that have similar VAT rules (e.g., Turkey and Norway) and in the tradable services sector (rather than goods and services), due to VAT procedures undertaken at customs (e.g., carbon credits and mobile phone minutes).[27]

Contra-trading. Contra-trading schemes are one of the more complex types of MTIC fraud. In addition to the transactions taking place under a simple carousel fraud scheme, here fraudsters introduce both legitimate and fraudulent transaction chains, in parallel to each other. With this approach they are able to further: (1) hinder the tax administration's detection of fraudulent activity, and (2) allow the contra trader to minimise its VAT liabilities.[28] The existence of a contra trader that acts in a legitimate manner makes it harder to detect the fraudulent activity, thereby extending the lifespan of the scheme.[29]

Cross invoicing. The inclusion of cross-invoicing is another strategy used to delay detection and reduce the VAT liabilities of the missing trader/cross-invoicer. With this approach the missing trader does not disappear immediately and, rather than using a legitimate transaction chain to hide the fraud, incorporates fictitious invoices. Cross-invoicing schemes involve a fabrication of invoice chains that either do not correspond to the actual movement of goods or are used to offset VAT liabilities incurred by the missing trader.[30] The flow of goods and services in this scheme, as well as the movement of invoices, bears certain characteristics of a contra-trading scheme; however, it uses a chain of fictitious invoices – rather than a parallel legitimate chain of transactions – to avoid detection.[31]

Customs procedure 42. According to Customs Procedure 42, businesses can request a VAT exemption if they are importing goods from outside the EU, but for final consumption in another Member State.[32] In some instances, MTIC fraud schemes also involve parties from countries outside the EU, and can potentially trigger Customs Procedure 42. Customs Procedure 42 fraud occurs when businesses apply for a VAT exemption when importing goods from outside the EU, but instead of being moved to the destination country (where they can be taxed) the goods are released domestically for consumption. Incorporating elements of Customs Procedure 42 fraud in MTIC schemes makes it increasingly difficult to detect the fraud.[33]

Other approaches used to avoid detection. As tax administrations implement techniques to clamp down on VAT fraud (see Box 1), a range of other techniques used to extract the maximum value from a scheme have appeared. For example, to avoid increased scrutiny when registering for VAT, fraudsters might use dormant, rather than newly registered, companies as the missing trader, or register the business in a sector which is under less scrutiny from tax administrations. In some cases, there are

---

[26] See: Szabo (2019).
[27] See: Ainsworth (2010).
[28] For details on the contra trader fraud, see HMRC (2022), HMRC Internal manual. VAT Fraud. VATF23550.
[29] See Appendix A for a detailed analysis of the contra trader scheme.
[30] Based on: FISCALIS (2018) and Szabó (2019).
[31] See Appendix A for a detailed analysis of the cross-invoicing scheme.
[32] See: https://www.asd-int.com/en/what-is-a-customs-regime-the-example-of-the-customs-regime-42/.
[33] See: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/731902/IPOL_STU(2022)731902_EN.pdf.

reports that trusted individuals are used to help decrease the due diligence that legitimate companies undertake on new purchases.[34]

Overall, compared to some other forms of VAT fraud (e.g., knowingly and incorrectly charging the zero rate rather than the standard rate on domestic transactions), MTIC fraud is particularly sophisticated. It requires an understanding of the VAT repayment system and, increasingly, of how to ensure that the missing trader scheme is set up in a way that avoids detection. As such, it is commonly carried out by highly skilled organised crime groups, with vast knowledge and resources at their disposal, allowing them to put in place highly complex structures that are distinct from one another (see Box 1). This diversity in the exact structure of individual frauds means that developing a precise typology is less helpful than focusing on the common features across schemes, as discussed in the next section.

I.c. Measures to tackle MTIC fraud

The measures implemented by Member States and the EU to tackle MTIC fraud (see Box 1) and the timing of this implementation are important for the estimation of MTIC fraud. Some of those measures hinder the operationalisation of fraudulent schemes, whereas others, such as the domestic reverse charge, are expected to eliminate MTIC fraud for groups of goods and services and Member States concerned. Tracking the discontinuity of suspected indicators of fraud, discussed in Chapter IV, around the dates of the implementation of such mechanisms could be one of the means to test the suitability of the proposed methodological approaches.

Box 1: Measures to combat VAT fraud

Domestic reverse charge. Since 2007, Member States have been able to implement the reverse charge on domestic transactions under certain circumstances, pursuant to Council Directive 2006/112/EC of 28 November 2006,[35] with Council Directive (EU) 2022/890 of 3 June 2022[36] extending this temporary provision until 31 December 2026. The circumstances under which the reverse charge can be applied by a Member State cover a number of goods and services that have been the subject of large MTIC frauds, such as mobile phones or raw and semi-finished metals. Implementing the reverse charge in these areas effectively closes the door to MTIC fraud: where the reverse charge is in place, the purchaser must account for and remit output VAT on the acquisition, which deprives the missing trader of the opportunity to charge VAT and disappear without remitting it to the tax administration. There is evidence that these regimes are effective,[37] which in fact forces fraudsters to move their operations to other goods and services not covered by the reverse charge mechanism, or to Member States that have not chosen to apply the domestic reverse charge on those goods.

More stringent registration procedures. Part of every MTIC scheme is the disappearance of a trader. This can be easily done by setting up a new company, as no time or effort is needed to trade or purchase assets before engaging in the fraud. Over time, Member States have put in place measures to deter missing traders from setting up new companies. Some of the less drastic measures include increasing the registration thresholds, subjecting new registrations in sectors where frauds are common to higher scrutiny (however, as noted above, this may

[34] Based on experience of VAT practitioners from across the EU.

[35] See: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02006L0112-20070101.

[36] See: https://eur-lex.europa.eu/eli/dir/2022/890/oj.

[37] A recent study estimated that the introduction of the domestic reverse charge mechanism in Germany has inhibited the volume of MTIC fraud amounting to approximately 5% of VAT revenues between 2009 and 2018. See: Buettner, T. & Tassi, A. (2023). VAT fraud and reverse charge: empirical evidence from VAT return data. International Tax and Public Finance. Available at: https://link.springer.com/article/10.1007/s10797-023-09776-y#Abs1.

simply lead fraudsters to set them up in a different sector), or generally introducing more extensive registration procedures.[38]

Due diligence and penalties for involvement in VAT fraud. MTIC schemes often rely on the involvement of legitimate companies to perpetuate the fraud. However, it is often not clear whether a company involved in a chain of fraud is truly "innocent", or in fact a company with correct compliance, which is nevertheless complicit in the fraud. Litigation can often hinge on whether the company involved should have known, or conducted sufficiently rigorous due diligence to have known, about the fraud. Poland is one example where a set of such measures was implemented to reduce the overall VAT compliance gap.[39] The tax administration applies basic Know Your Customer procedures before proceeding with transactions, including: verification of VAT status on the Ministry of Finance website, verification of VAT EU number in VIES, verification of company details in the National Court Register and the Central Register and Information on Economic Activity. Inadvertent participation in carousel fraud due to lack of appropriate due diligence procedures and the use of fictitious invoices is punishable by up to 5 years' imprisonment and financial penalties. In cases where the value of falsified invoices exceeds PLN 10 million, the taxpayer can be imprisoned for up to 25 years.[40]

Tracking of products. Tracking of specific products can help limit the cost of fraud by helping Member States identify that certain products are being cycled through a carousel scheme. This has been done in the case of mobile phones using the IMEI number, for instance.[41] Similarly, real-time monitoring of the movement of goods vulnerable to tax fraud was also implemented in Romania in January 2023. The Romanian e-Transport System requires businesses to register a journey in advance with the tax administration and fit transporting vehicles with tracking devices.[42]

E-invoicing. E-invoicing involves counterparties to a transaction reporting their sales and purchases electronically to the tax administration, ideally in real time. Based on the nature of VAT (with both input and output VAT typically being recorded and reported by two separate entities), tax administrations can use these invoices to check if output VAT has been declared in cases where input VAT is being reclaimed. As missing traders are less likely to supply such invoices and do not remit the VAT (which is the essence of the fraud), e-invoicing can help Member States detect MTIC fraud. However, the ability to limit the loss of revenue is contingent on how fast e-invoices are reported and VAT is collected, and historically has been limited, as intra-Community invoices do not always fall under Member States' schemes. Thus, e-invoicing implementation is sometimes paired with real-time reporting obligations, or pre-clearance mechanisms. For example, Italy's Sistema di Interscambio makes e-invoices available to trading parties and the tax administration in real time. Additionally, VAT deductions are only allowed if the invoices are correctly processed within an appropriate time frame.[43] The upcoming VAT in Digital Age reform package also aims to introduce standardised real-time digital reporting using e-invoicing.[44]

Split payments. Split payment is a mechanism under which the payment for goods or services (net of VAT) is made by the purchasing entity and credited to the supplier's usual bank account; however, the VAT due is paid to a separate account. This separate account into which the VAT is deposited is created automatically by the bank

[38] For instance, in Spain, businesses need to register for EU VAT through a special form, which can only be submitted via a secure online portal or filed in person at a local tax administration office. See: Carbonell J. (2022). Form 036 and 037: registering with the tax authorities. Tas Consultoria.

[39] Podatki.gov.pl (2020). Wykaz podatników VAT. Pytania i odpowiedzi. Ministry of Finance.

[40] Bartczak A. (2021). Nie daj się wkręcić w karuzelę VAT!. Księgowość jest sexy. Grant Thornton.

[41] Nemesis, a system implemented in 2006 by UK's HMRC, which tracked exports of mobile phones using each device's IMEI identification numbers. Shipments of phones had to be stamped and sealed by customs officers at the point of exit, and exporters had an obligation to record IMEI numbers of devices involved in a central database. Pollock I. (2006). The Nemesis for VAT fraudsters? BBC News.

[42] Storecove (2023). Implementation of RO e-transport system in Romania.

[43] Mandatory for the majority of B2B, B2C, B2G domestic and intra-Community transactions. See: Avalara. Italy Sistema di Interscambio real-time e-invoices.

[44] See: https://taxation-customs.ec.europa.eu/taxation-1/value-added-tax-vat/vat-digital-age_en.

and access to it is restricted, as any balance on it can only be used to settle tax liabilities with the national tax administration.[45] The only example of an extensive and successful implementation of this measure is the Polish split-payment mechanism, introduced in November 2019, mandatory for domestic B2B transactions involving goods and services vulnerable to VAT fraud and with the invoiced value exceeding PLN 15 000.[46] Other examples include a partial and temporary implementation in Italy,[47] and a halted introduction in Romania.[48]

SAF-T reporting. SAF-T reporting is a measure implementing an OECD-designed Standard Audit File for Tax (SAF-T). The file takes the form of a harmonised XML file containing company transaction data[49] submitted to the tax authority in place of a paper-based VAT return. The structure and file syntax are uniform across jurisdictions that have adopted it, enabling the implementation of automation in data exchange and analysis. As of February 2022, seven Member States had introduced SAF-T reporting in a mandatory or on-demand form for all VAT taxpayers.[50]

EU-level measures. Over the years, measures have been put in place at an EU level to help combat MTIC fraud. One of the most prominent measures were implement in consequence of adopting The Council Regulation No. 904/2010[51] on administrative cooperation and combating fraud in the field of value added tax provided the basis for the creation of the **EUROFISC network** in 2010. This network aims to combat cross-border VAT fraud, and consists of liaison officials from the EU-27 and Norway who are mandated to work on coordinated access, the processing and analysis of data, planning of joint action, and direct cooperation with EUROPOL and the European Anti-Fraud Office. Based on insights from their analysis, and available information and analyses, the officials can also take action at a national level, including information, audit, and VAT number deregistration requests.[52] In 2019, the EUROFISC network began utilising the Transaction Network Analysis (TNA) tool to leverage data mining techniques to detect suspicious activity.[53]

The **VAT Information Exchange System (VIES)** is another EU tool that helps tackle MTIC fraud. VIES is primarily a platform for information exchange between Member States' tax administrations. On top of this, VIES allows companies to verify if the businesses they are working with are VAT registered and allowed to engage in intra-community trade.[54] With the introduction of the EU VAT quick fixes in 2020, businesses are required to validate the VAT registration number of the acquirer in order for the Intra-Community Supply to qualify for zero rated VAT. Businesses can use VIES to validate the VAT registration details of the acquirer.[55]

Cross-border exchange of information was also an aim of the creation of the European Public Prosecutor's Office (EPPO). EPPO is a legal body that was established in 2017 under the Treaty of Lisbon provisions; it began its

[45] KPMG (2019). Split Payment Mechanism: a controversial tool for fighting VAT fraud.
[46] KPMG (2019). Split Payment Mechanism: a controversial tool for fighting VAT fraud.
[47] Mandatory for transactions involving payments made to public authorities, or involving companies controlled by central and local public authorities, and companies listed on the Milan Stock Exchange. See: Marosa VAT. Italy extends VAT split payments regime.
[48] A split payment mechanism was introduced in 2018, but was scrapped in 2020 after the EU Commission noted its incompatibility with the EU VAT Directive and that it was a disproportionate burden for honest entities. It was mandated for all insolvent or VAT indebted suppliers, and available on voluntary for other businesses, incentivised with a 5% discount off corporate income tax liabilities. It now remains optional on a voluntary basis. See: Asquith R. (2020). Romania withdraws VAT split payment. Avalara.
[49] This can include account books, bank statements, stock warehouse/storage, VAT sales and purchase ledgers, VAT invoices, Revenue and expense tax books, and Income records. See: Avalara. Poland SAF-T.
[50] VATCalc (2023). Which countries have introduced OECD's SAF-T.
[51] See: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32010R0904.
[52] See: https://taxation-customs.ec.europa.eu/taxation-1/vat-and-administrative-cooperation_en.
[53] The tool utilises "automated data mining", connecting contents of VAT returns and compliance data collected on IT platforms of national tax authorities. TNA allows Eurofisc's anti-fraud experts to monitor discrepancies in disclosed data and detect suspicious behaviour. See: European Commission (2019). VAT Fraud: New tool to help EU countries crack down on criminals and recoup billions.
[54] See: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/731902/IPOL_STU(2022)731902_EN.pdf.
[55] See: https://www.pwc.nl/en/insights-and-publications/tax-news/pwc-special-budget-day/2020-tax-plan-vat-quick-fixes.html.

operations in 2021.[56] Another example of institutional cross-border cooperation is to be found in the creation of the Transaction Network Analysis (TNA) tool in 2019 to leverage data mining techniques to detect suspicious activity.[57]

Looking forward, the Commission has announced its **VAT in the Digital Age (ViDA)** proposals, aimed at amending the rules governing VAT obligations in the context of intra-Community trade. The main three objectives of the reform package are: (1) the introduction of standardised Digital Reporting Requirements (DRR), allowing national tax authorities to monitor cross-border transactions through real-time e-invoicing; (2) updating existing VAT rules to reflect market developments in platform economy activities; and (3) improving the OSS and IOSS regimes by adjusting the reverse charge mechanism and introducing a single VAT registration.[58] If implemented, the first objective, introduced in stages between 2024 and 2028, would apply anti-fraud measures already present in some Member States to the entirety of the EU, potentially improving the detection and prevention of MTIC fraud.[59]

I.d. Commonalities across different MTIC schemes

While there is a lot of diversity in the exact structure of MTIC schemes, there are also some key common features, which are relevant for defining the scope of the calculation and for assessing different approaches to estimating the size of the MTIC fraud gap. These key commonalities, laid out in Table 1, summarise the analysis of the characteristics of MTIC fraud and serve as an extension of the short definition of MTIC fraud provided in the introduction.

Table 1: Key commonalities across MTIC schemes

Commonality	Description	Assessment/implications
Intra-Community Supply / Acquisition has occurred	Whether on paper or in reality, a supply has crossed between Member States in a transaction.	Trade statistics – in particular Intrastat, but also data gathered through VAT returns and EC sales list returns – are an important source of information that could potentially be used to estimate the scale of fraud, even if in some schemes there is no physical movement of goods involved.
The fraud relies on the missing trader taking advantage of the zero rating of Intra-Community Supply	What distinguishes MTIC fraud with an intra-Community element from domestic fraud is that the missing trader takes advantage of the intra-Community rule of zero-rated supply transactions that create opportunity of high profit margins from fraud. The fraudsters also benefit from slower or lack of information exchange across the border.	

[56] EPPO's jurisdiction spans the participating Member States, and it is within its power to investigate and prosecute crimes against the Budget of the European Union, embezzlement of EU funds, and intra-Community VAT fraud cases, as long as estimated damages exceed EUR 10 million. European Council (2022). European Public Prosecutor's Office. Available at: https://www.consilium.europa.eu/en/policies/eppo/#:~:text=The%20European%20Public%20Prosecutor's%20Office%20(EPPO)%20is%20an%20independent%20body,corruption.

[57] The tool utilises "automated data mining", connecting contents of VAT returns and compliance data collected on IT platforms of national tax authorities. TNA allows Eurofisc's anti-fraud experts to monitor discrepancies in disclosed data and detect suspicious behaviour. European Commission (2019). VAT Fraud: New tool to help EU countries crack down on criminals and recoup billions.

[58] PricewaterhouseCoopers (2022). VAT in the Digital Age proposals published by the European Commission. Tax Policy Alert.

[59] PricewaterhouseCoopers (2022). VAT in the Digital Age proposals published by the European Commission. Tax Policy Alert.

Commonality	Description	Assessment/implications
Missing traders face mixed incentives to file returns and declarations	The behaviour of fraudsters depends on their expectation of whether the discrepancy in mirror registers or the reporting of high-value Intra-Community Acquisition is more likely to alert the administration and trigger an audit. This expectation had likely been driven by the past actions of administrations and the evolution of means to exchange of information and track fraud.	The methods based on tracking discrepancies in the mirror registers could be used to estimate the scale of MTIC fraud. However, these figures need to be treated with caution as some missing traders may duly report their transaction (in their tax return and Intrastat).
As opposed to the missing trader, other companies involved in the fraud have an incentive to file returns	In order to recover input VAT, there is a monetary incentive to maintain records and file returns. If other parties in the transaction chain are complicit in the scheme, correct filing also helps maintain a cloak of legitimacy.	The ability to verify the other side of a transaction is central to many Member States' approaches to tackling VAT fraud (e.g., e-invoicing), and could be relevant for MTIC.[60] It may also suggest that not only tax returns, but also Intrastat returns, would be filed by non-missing traders to demonstrate compliance.
High transaction values	MTIC schemes are typically high values and quickly executed. This leads to large upticks in traded volumes of specific supplies.[61]	For large schemes, it is possible that these upticks are visible in periodic (e.g., monthly) data on purchases or sales of particular goods and services in companies' internal data. As aggregated turnover data is not typically at product level, such upticks may not be visible in aggregated data on supplies. Trade data at product level could reveal these upticks, but relies on the correct classification of products, and sufficient scale of the scheme. In the case of carousel and other frauds, where good and services are not consumed in the destination but are further re-exported, upticks in trade volumes could be expected both for Intra-Community Acquisition and Sales (if reported correctly, see row above).
Certain types of goods and services are more susceptible to MTIC fraud	Supplies that maximise profits (by maximising revenue and/or minimising cost), may be more commonly the subject of fraud, although there is no exhaustive	There is potential to focus the analysis on a set of goods and services that exhibit certain characteristics; however, such an analysis would not be exhaustive. Many Member States have

[60] ViDA's EU-level e-invoicing and real time reporting provisions are of particular importance here.

[61] For instance, in the case of a precious metal company, which was affected by a missing trader fraud scheme, a senior director of the company was the first to realise the fraud was taking place when by chance he observed the large amount of the precious metal at issue in their vault, when previously the company had only been dealing with very minimal quantities of the metal.

Commonality	Description	Assessment/implications
	list (see in the running text below).	implemented a reverse charge mechanism on such goods/services or have an idea of which goods have recently been at the centre of MTIC schemes, giving an initial but not exhaustive indication of where MTIC fraud may lie.
Fraudsters react to measures hindering the fraud and move to different markets and jurisdictions	There are reports of Member States having reduced their VAT losses after implementing measures such as the domestic reverse charge, e-invoicing, split-payments, etc.	The introduction (or introduction and subsequent withdrawal) of such measures could potentially be used to help identify the scale of VAT fraud in certain cases. However, such data may not identify MTIC fraud uniquely. Moreover, in some cases, the measures introduced (and later reversed) were onerous for businesses, which may mean that any reporting following such measures would not be reliable.

Source: own elaboration.

As noted in Table 1, the goods and services typically subjected to MTIC fraud are those which enable fraudsters to maximise profits (by complicating detection, minimising costs, increasing profit margin or through lowering market entry barriers). The categories of goods and services used by fraudsters are:

- High-value low-volume goods (e.g. mobile phones, electronics, precious metals), allowing quick operations involving large gains.
- Goods taxed at standard rates to increase the margin of fraudsters' earnings.
- Goods for consumption (e.g. foodstuffs, such as candy, grain, oils; fuels), so that goods disappear from the market.
- Goods for which it is common practice to sell at a loss in certain periods (e.g. mobile phones), to prolong the moment of detection.[62]
- Services and intangibles (carbon credits (historically), telephone minutes, energy), to avoid transportation costs.

Revealed cases of fraud enumerated in Appendix G confirm the prevalence of such goods and services in fraudulent schemes. They also point to the scale of forgone revenue on per-case basis. According to the releases, forgone VAT revenue caused by the criminal groups targeting specific country and groups of product ranges typically from several to several dozen EUR million. This means that trade value of products in MTIC schemes run by a single group could range several hundred EUR million – non-negligible values when compared to licit transactions. As a result, fraudulent transactions should be expected to have a significant impact on the statistics described in more details in Section IV.a.

Over the years, we have witnessed the movement of fraud between jurisdictions, as well as an evolution of fraudulent schemes as the EU and Member States put in place measures to combat specific

[62] For instance, where items are purchased in bulk for a launch or seasonal occasion, and the remaining units are resold and heavily discounted.

aspects of fraud (see Box 1), but relatively little change in the overall goods and services involved in MTIC fraud. Some schemes have closed: most notably the carbon credits fraud, which was brought to an end once the credits were reclassified as exempt. However, an early case of MTIC fraud is related to mobile phones; and despite awareness and Member States' countermeasures, mobile phones and electronics continue to be the subject of large MTIC schemes.[63]

The aim of the overview presented in this chapter was to highlight the complexity MTIC fraud in practice. The schemes used by fraudsters build on several main theoretical "archetypes". These archetypes vary in terms of complexity and strategies employed to delay detection, but at the same time share many commonalities, which are important to keep in mind when developing a common approach to estimating the size of the MTIC fraud gap. Inherent to MTIC fraud is the existence of an intra-Community transaction, the appearance of large upticks in trade volumes and varying susceptibility of different goods and services to this kind of fraud. Furthermore, in the EU context, missing traders exploit the Intra-Community rule of zero-rated supply transactions, which makes the fraud more lucrative.

Although the general pathways of fraud are known, revealed cases of MTIC fraud demonstrate that, in practice, fraudsters' approaches are ever-changing and include various additional elements and techniques, designed to further increase the time and resources needed to uncover them. This makes MTIC fraud detection particularly challenging, often necessitating the use of novel approaches and international cooperation. In order to aid these efforts, several EU-level measures aiding the exchange of information between Member States, such as the VIES or the introduction of more stringent and uniform reporting requirements, have been introduced. In addition to these detection efforts, both individual Member States and the EU as a whole have introduced and make use of a wide range of measures to combat fraud (both MTIC fraud and VAT fraud more generally), such as split payments.

II. Review of earlier work

This chapter forms the foundation and necessary first step towards defining the methodological options and identifying methodological gaps. It takes stock of the work carried out up to date by reviewing academic literature and reports in the public domain and unpublished materials from the Fiscalis Tax Gap Project Group[64] (see Section II.a). This review of scientific articles, working papers and reports of international organisations, materials published by national tax administrations and national statistical institutions was needed in order to systematize the analytical methods that are available to researchers and administrations (see Section II.b). Based on this mapping, supplemented by a survey and in-depth interviews with Member State administrations, the study team gained a comprehensive view of the methods and data sources that might be useful for estimating the scale of MTIC fraud (see Section II.c). The findings from empirical papers summarised in this chapter are also the reference point for the estimates presented in Chapter VII.

II.a. Literature review

The literature identified during the preliminary literature review (which totalled 25 articles and reports) was grouped by the degree of relevance to our study (see Table 2). Given the goals of this study, the

[63] Compare: Europol (2021). VAT fraud clampdown: international scam with memory cards uncovered in the Netherlands.; and Eurojust (2019). Successful action against VAT fraud with mobile phones.

[64] The group was established under the FISCALIS 2020 programme, with the goal of pooling knowledge and providing a space for sharing experiences from already carried out tax gap estimations. The group consists of national experts from Member States and its work is coordinated by the Directorate General for Taxation and Customs Union (DG TAXUD) of the European Commission.

aim was to first and foremost identify papers which touched on the methodological approaches to MTIC gap estimation used up to date and, ideally, presented their own estimates. Upon examination of the methodological and empirical literature, primarily containing empirical studies and methodologies for estimating the MTIC gap,[65] the study team established that the methodologies used to date could be divided into three categories: empirical studies, extrapolation from direct approach estimates (either of the size of the MTIC gap or its share in the VAT gap) and expert qualitative assessments.

Overall, we found that work which both offered estimates and was transparent with regards to the specific methodology applied, was scarce – only nine of the identified sources met these criteria (among them were empirical studies presenting own MTIC gap estimates, summary documents such as reports from the European Commission and literature presenting methodologies). Among the methods used in these papers were the analysis of discrepancies in bilateral trade statistics, econometric modelling, application of option pricing models, time series analysis of VAT revenues and repayments and econometric forensic methods. In addition to this, several sources offering estimates but keeping the methodology confidential were identified, chief among them the work carried out by Her Majesty's Revenue & Customs, whose estimates are considered to be among the most reliable, as they were sourced from operational data (Vaškovič, Zídková & Arltová, 2021). As a result, although they do not contribute to the mapping of methodologies, they do nevertheless offer a valuable point of reference.

Another group of papers offering estimates were sources which based their calculations on the extrapolation of already existing estimates. However, by their very nature, studies of this kind are not useful for the task of methodological mapping and work on the strong assumption that the original estimates are reliable and can be applied indiscriminately, regardless of context (e.g. to a different Member State).

The most numerous group of sources identified were those which did not contribute to knowledge on the methodological approaches used up to date or provide own estimates, but instead provided valuable insights on different aspects of MTIC fraud, relevant to this study to varying degrees. Some examples were papers consisting of in-depth discussions on this type of fraud and its consequences, providing detailed information on the sectors and products frequently targeted by fraudsters, summarizing the steps taken to date to combat MTIC fraud (both on a national and EU level), providing overviews of the fraud schemes employed, insights from specific cases of MTIC fraud uncovered by Member States' tax authorities and analyses from various perspectives (e.g., Sokanovic (2017), where MTIC fraud is discussed through the lens of criminal law). Taken together, this information drew attention to some important considerations, potential areas of focus, things to keep in mind when developing the methodological approach. For instance, by uncovering the most common targets of this type of fraud, the study team was in a better position to determine which products and sectors deserved particular attention when testing for the presence of fraud. Information on the specific schemes used, especially more recent cases, allowed the study team to better consider which methodological approaches would be best suited to capturing a variety of fraud schemes. The grouping of identified sources is presented in Table 2.

[65] The Team also included literature which covered all VAT fraud but could still be seen as dealing primarily with MTIC fraud (e.g., Frunza, Guegan & Lassoudière, 2010).

Table 2: Classification of articles and reports

			Number of papers	Examples
Literature on MTIC fraud	Presenting estimates of the scale of MTIC fraud	Empirical studies	7	*CASE (2015)*
		Studies extrapolating other estimates	1	*Borselli (2017)*
		Expert qualitative assessments	1	*EY (2015)*
	Other	Summary documents (European Commission reports)	1	*Fiscalis (2018)*
		Methodologies for estimating the MTIC gap	1	*David and Semerád (2014)*
		Methodologies for detecting MTIC fraud	1	*TARC (2020)*
		Other (e.g., its impact, mechanisms to reduce scale)	24	*Ainsworth (2011)*
		Deemed of low relevance	5	*Agha & Haughton (1996)*
General methodological literature			4	*Chow et al. (2010)*

(turquoise)	of primary interest
(light turquoise)	of secondary interest
(pink)	excluded from further analysis

Source: own elaboration.

Below we present short summaries of studies presenting original estimates of the scale of MTIC fraud, with an emphasis on empirical approaches using a variety of top-down methodologies. High representation of methods employing a top-down approach could be justifies by the fact that, bottom-up approaches using individual-level data are rarely available for researchers.

Braml & Felbermayr (2021)

The authors of this paper take on the problem of EU self-surplus – i.e. the reported values of exports exceeding the reported values of imports. Using forensic accounting methods, they find that VAT fraud is a key driver behind this surplus, particularly in cases of neighbouring countries with differences in applied VAT rates. Using Eurostat data on bilateral trade flows they estimate the export biases for each EU country, by taking the mean discrepancies from each country pair. They then correlate their measure of country discrepancies with VAT compliance gaps, estimated by Morrow et al. (2019) and find significantly positive correlations between the two in the case of goods discrepancies (but none for services, which could reflect that they are less affected by VAT fraud). Ultimately, the authors identify countries with the most accurate statistical regimes at the time of the study (Cyprus, Ireland, Luxembourg, Sweden, and the Netherlands), identify the e-commerce market as one of the causes of the continually growing trade discrepancies and value the EU-wide losses from VAT fraud at EUR 27-35 bn annually, which rises to EUR 64 bn in the worst-case scenario.

CASE (2015)

This study offers an example of estimates based on algebraic operations and using trade mirror statistics. The authors used Intrastat trade data for 2014 and the first half of 2015 to determine the reduction in trade size discrepancies between Polish companies and their European partners following the introduction of the Reverse Charge Mechanism (RCM) on the Polish steel product market. They arrived at a figure of PLN 537 million (ca. EUR 128 million) annually, which was a proxy for the value of MTIC fraud in this sector. Of note is the fact that the actual reduction in losses incurred by the Polish government amounted to PLN 424 million, as the introduction of the Reverse Charge Mechanism led to an increase in MTIC fraud by an estimated PLN 113 million on other markets.

Ernst & Young (2015)

Ernst & Young (2015) use a qualitative assessment to estimate the size of both VAT fraud, and MTIC fraud in particular, in the context of the proposal by the European Commission to introduce the destination principle to intra-Community trade. For this purpose, the authors sent surveys to tax authorities across all EU-28 Member States,[66] in which they asked them to provide their own assessments of the share of the VAT compliance gap arising from intra-EU B2B trade in their respective countries that could be attributed to fraud and, furthermore, how much of it was caused by MTIC fraud specifically. Unfortunately, only nine Member States responded to the survey, with one Member State failing to provide an estimate of the share of MTIC fraud.[67] The respondents estimated that, on average, 36% of the VAT compliance gap in their respective countries was due to VAT fraud. On average, they considered 20% of the overall VAT gap to be due to MTIC fraud, with a weighted average (weights based on overall VAT compliance gap proportion) of 24%. If applied to the 2011 estimate of the overall EU VAT gap,[68] which amounted to EUR 193 billion, this would result in an overall EU MTIC fraud estimate of EUR 46.3 billion in lost revenue due to MTIC fraud alone (for the 26 Member States and using the weighted average estimate).

Frunza (2016)

Frunza (2016) used Eurostat data on VAT revenues and Intra-EU trade gaps, imports, and exports (for the years 1999-2014) and employed a combination of econometric and option pricing methods. The reasoning followed that any increase in EU imports in a given country should result in a proportional increase in VAT collected (theoretical VAT) and that unexplained increases in imports, above the normal economic level linked to real demand, are likely to be the result of MTIC fraud. Thus, the VAT compliance gap formed due to MTIC fraud could be expressed as proportional to the difference between actual imports and imports at the normal level. As noted by the authors, the resulting formula resembled a known equation for the value of a financial instrument known as the vanilla call option. Taking advantage of this, the authors modelled the VAT compliance gap of a given country as a vanilla call on the value of imports in that country, capturing the sensitivity of collected VAT with respect to imports and the overall size of the MTIC gap.

Gajewski and Joński (2022)

The authors estimate the MTIC gap in Poland looking at VAT revenue and VAT repayments. Their main assumption is that in “normal” conditions, the VAT collected and reclaimed should move together.

---

[66] Including the UK.

[67] Those nine Member States were: Austria, Bulgaria, Czechia, Cyprus, Finland, France, Slovakia, Slovenia and the UK.

[68] Based on 26 out of the EU-28 Member States. The VAT compliance gap is defined as the difference between the theoretical VAT liability and the actual VAT collected.

However, organized VAT fraud, such as MTIC fraud, increases the ratio of VAT reclaims to VAT collected, lowering overall VAT revenues. This sets it apart from growth in unreported (shadow) economic activity, conducted outside the official economy, is likely to reduce both VAT collected and reclaimed, leading to a decline in overall VAT revenues. The proposed approach, applied for analysing the VAT compliance gap in Poland, uses the evolution of three time series: VAT collected, reclaimed, and total revenues. The VAT collected and reclaimed closely tracked one another from 2008–2011, then diverged as VAT reclaims rose significantly higher than VAT collected, decreasing total VAT revenues. This trend aligns with increased VAT fraud during that time. The two series re-converged in 2017–2018, reflecting successful measures against organised crime, particularly carousel fraud, before diverging slightly again in 2019. To further understand the fraud's contribution to the VAT compliance gap, the study looked at the difference between the actual VAT repayments and the theoretical reclaims derived from the 2008–2011 ratio of VAT repayments to VAT collected. This difference represents the fraction of the VAT compliance gap attributable to unexplained reclaims, including fraud. Subtracting this from the total VAT compliance gap allowed for a breakdown into two components: the portion attributable to unexplained reclaims, indicative of fraud, and the portion due to other factors such as the shadow economy, bankruptcies, and honest errors.

<u>David and Semerád (2014)</u>

David and Semerád (2014) offer an approach to calculating VAT evasion based on the Czech fuel market. To achieve this, the authors combine data on 2012 distributor prices and data from surveys of interested entities (154 responses overall), which asked them for estimates of the difference between the prices of usual suppliers and those potentially affected by fraud. In the next step, they delineate the price "risk zone" – prices which are below those attainable without illegal action, and thus potentially linked to fraud. Ultimately, they express the potential tax evasion amount as the product of the nominal VAT rate and the sum of the products of the quantities and their assigned prices reported by distributors, limited to only those where the price is in the risk zone.

<u>Frunza, Guegan, & Lassoudière (2010)</u>

In this study, forensic econometric methods are used to determine the extent of VAT fraud on the European carbon allowances markets between the end of 2008 and beginning of 2009. The authors build on a previous work in which they demonstrated that Arbitrage Pricing Theory-like models can be used for quantifying the impact of individual factors such as coal or energy on the market. More specifically, they found that these factors explain over 75% of the behaviour of CO2 prices. They observed a drop in the performance of their model at the end of 2008, the alleged start of the carbon emissions allowance carousel scheme, followed by a rapid improvement in response to an anti-fraud legislation. The estimated extent of MTIC fraud on the carbon market was estimated at EUR 1.3 bn.

<u>Vaškovič, Zídková and Arltová (2021)</u>

In this paper, the authors estimate the volume of MTIC fraud between Poland and Czechia in on the electronic device market, using an ex-post analysis of trade balances between the two countries. To do this, the authors use Eurostat data on international trade and assume that all MTIC fraud in this market is eliminated upon the introduction of the specific reverse charge mechanism for these goods. The authors used econometric regression to confirm that MTIC fraud was taking place and simple algebraic operations to calculate the gap itself. Overall, they found that, under their assumptions, MTIC fraud on this market was responsible for the loss of EUR 44-51 million in VAT revenues on the Czech side in 2014 and the 1st quarter of 2015.

Fiscalis 2018

The Fiscalis (2018) report does not offer its own methodologies, but instead is an exhaustive resource on VAT fraud, and MTIC fraud in particular. It contains information on MTIC schemes and a thorough list and comparison of existing top-down and bottom-up methodologies (both for MTIC fraud detection and estimation), along with a discussion on available data sources.

The available estimates of the MTIC fraud, presented in Table 3 below, enable cross-validation of the overall scale of the forgone revenue due to MTIC fraud in the EU. Unfortunately, the evidence is too scarce and too scattered to compare the estimates for particular Member States, groups of products, or time periods. This could not only give insights on the scale of specific irregularities but also on the accuracy of different methods employed in the literature. Overall, the estimates in the literature are internally consistent (see Table 3). The EU-wide forgone revenue in recent years varied across years and studies from EUR 27 to 93.5 billion. Such a large scale was in line with high estimates of forgone revenue from MTIC fraud in particular countries and certain goods and services.

Table 3: Comparison of MTIC gap estimates across empirical studies

	Ruffles et al. (2003)	Frunza et al. (2010)	CASE (2015)	Frunza (2016)	Braml & Felbermayr (2021)	Čejková & Zídková (2019)	HMRC (2019)	Vaskovic et al. (2021)
Member State covered	UK	EU-wide	Poland	EU-wide and per EU MS	EU-wide	Czech Republic	UK	Czech Republic
Time covered	1999-2002	2008-2009	10.2012-09.2013	2013-2014	2006-2018	04.2010-03.2011	2005-2017	1Q 2014-1Q 2015
Products/ services	all goods	carbon market	steel products	all	all	waste and scrap	all	electronic devices
Estimation (revenue forgone)	ca. EUR 15 billon (2002)	EUR 1.3 billion	ca. EUR 130 million	EUR 82.5 billion (2013) EUR 93.5 billion (2014)	EUR 27-64 billion per year	EUR 56 million	ca. EUR 3.6-5 billion (2005-6) ca. EUR 0.6 billion (2016-17)	EUR 44-51 million

Source: own elaboration

In addition to the methodologies applied to MTIC fraud gap estimation discussed above the team also identified literature concerning methodologies that could be employed for the purpose of such calculation, despite never having been applied to MTIC gap estimation itself. These methodologies have been added to the list of scenarios under consideration for Phase II of this study and are discussed in further detail in Section IV.c.

The scarcity of literature focusing specifically on estimating the scale of MTIC fraud and providing information on the methodological approach could be broadly attributed to two factors. First of all, uncovering this type of fraud and the production of reliable estimates of its scale is resource-intensive and involves many challenges. As discussed in Chapter 0, in practice the schemes used for MTIC fraud are ever-evolving and characterised by particular complexity and sophistication, with fraudsters employing various tactics to avoid detection and prolong the scheme's duration. What is more, fraudsters are able to rapidly switch between the sectors they target in response to new developments which affect the attractiveness and susceptibility to fraud of specific sectors or commodities. These shifts can be triggered by things such as the implementation of new legislation (for instance the introduction of a reverse charge mechanism in a specific sector), changes in commodity prices (e.g., in the energy market) or the introduction of new commodities particularly vulnerable to this type of fraud, for instance those involving no transportation costs (e.g., carbon allowances). These issues are further magnified by differences between Member States' data and approaches: gaps in the data and information exchange, differences in methodologies and definitions employed by individual Member States' tax authorities. As a result, what estimates do exist are bound to be imperfect, failing to capture the entirety of the fraud and lagging behind any new developments.

The second issue relates to the confidentiality of the methodologies behind some of the most often cited statistics. Estimates based entirely on operational data of national tax authorities (audits and tax returns), such as those of the UK HMRC or the Belgian Finance Ministry (HMRC, 2019; Reckon LLP, 2009), do not reveal the exact methodology behind them, often to the extent that not even an indication of the methodologies used is offered (Europol, 2018; EY, 2015). This makes it impossible to assess the methodologies and reliability of the estimate. The most prominent example of this is the widespread use of HMRC estimates of the share of MTIC fraud in the VAT gap (see for example Borselli, 2011, who arrived at a MTIC fraud estimate of between EUR 13 and EUR 23 billion).[69] The original direct estimates were limited to specific sectors (mobile phones and computer components)[70] and were ultimately abandoned in 2019, due to targeted measures against MTIC fraud lowering them to below GBP 0.5 billion (they can still be found in HMRC reports up to and including 2016-2017). Since the same cannot be said for MTIC fraud in other EU Member States, many studies extrapolate or refer to older HMRC estimates, preceding these targeted measures, rather than estimates corresponding to the same years. Thus, although extrapolation from existing estimates involves the least effort, it should be approached with a high degree of scepticism due to the aforementioned lack of transparency and the fact that extrapolating from national estimates in practice means that the results from one country (and usually also a specific timeframe and sector) are taken and indiscriminately applied to the whole of the EU, further decreasing their reliability.

However, it is important to note that some lessons can still be learned even from papers which offer little insight on their methodological approach. For instance, HMRC produces estimates covering the contribution of different types of fraudulent activities to total VAT fraud, given in the form of two values – the upper and lower bound of the amount of tax revenue lost. Initially the upper bound of the estimates was based on discrepancies between data on trade between the UK and EU Member States (upper because part of this discrepancy could be attributed to other factors) and the lower the application of a factor based on estimates from other countries. However, in 2006 HMRC changed its methodology in direct response to fraudsters' behaviour and began relying exclusively on what it called "operational evidence", declaring that estimates based on trade data were no longer reliable due to fraudsters' ever-changing tactics (Reckon LLP, 2009).[71]

---

[69] The estimate was calculated for EU-27, for 2009.

[70] A more detailed discussion of MTIC estimates produced by HMRC can be found in Ruffles et al. (2003).

[71] When discussing these estimates, it is also important to note that, following this change, the estimates produced covered all cases of attempted fraud, including the ones that were prevented (actual loss estimates were also provided). Moreover, the UK estimates treat trade statistics only as aggregate adjustments and relate to the value of missing trade transactions associated with MTIC (such as trade of phones and computer components), and not the value of the frauds themselves.

II.b. Mapping of methodologies

The analysis of articles and reports presenting estimates of the scale of MTIC fraud and related forgone revenue provide sufficient knowledge to classify the relevant analytical methods and describe their basic characteristics. Such an “inventory check”, presented in this section, was necessary for gathering more detailed information and the assessment of these methods.

The literature review shows a clear and strong distinction between the characteristics of the top-down and bottom-up methodologies employed. Top-down methodologies can be presented in a two-dimensional array, classifying the approaches by what variable the methods target (endogenous variables) and what type of quantitative methods are used. As depicted in Table 4, top-down methodologies either look at trade statistics or discrepancies in mirror registers, or analyse the VAT refunds. In most cases, these methods rely on specific observations of data patterns, which could be sufficiently described through simple algebraic operations. In only two of the analysed approaches did the authors use more complex methods, namely statistical methods or machine learning techniques.[72]

Methodologies based on discrepancies in trade statistics reported by exporting (Intra-Community Supply) and importing (Intra-Community Acquisition) partners appear to be of interest for this study. They are based on the observation that missing traders do not register fraudulent transactions in the Intrastat system. As differences in mirror statistics might be biased (e.g., due to different registration thresholds), the approaches employed could use statistical methods, such as k-means clustering,[73] to detect fraudulent deviations.[74] These methods could rely on very granular product categories (e.g., in the Combined Nomenclature (CN) or Harmonised System (HS) classification systems).[75] Other methods using trade statistics as a source of information are based on the assumption that quick shifts in trade volumes or large quantities of “risky” goods traded could stand for fraudulent transactions. The former regularity was observed by Čejková and Zídková (2019). The latter — by the *VAT gap in the EU* study.[76] In addition to the approaches based on trade statistics, methodologies estimating the value of legitimate VAT refunds utilise various economic aggregates and compare them with actual refunds.[77]

---

[72] Machine learning techniques are a broad group of algorithms that involve learning without explicitly being programmed (see Box 3).

[73] See Appendix A: Glossary of terms. K-means clustering is an unsupervised machine learning (see Box 3) algorithm which groups similar points in a dataset into clusters, with every point allocated to the with the nearest mean, with the goal of minimizing the within-cluster variance. As such, it is an exploratory data analysis technique which ensures that the data points within one cluster are as similar to each other and as dissimilar to points assigned to other clusters as possible. In the context of this study, this technique could be used to train an algorithm to classify each point in a mirror trade statistics dataset as fraudulent (bearing the characteristics of fraud occurring) or not, in the case of two clusters.

[74] See European Commission (2017).

[75] See CASE (2015) and Braml and Felbermayr (2021).

[76] See European Commission/CASE (2021). Not used for estimating the scale of MTIC fraud per se, but rather as a proxy to explain variation in the overall VAT compliance gap.

[77] See Gajewski and Joński (2022).

Table 4: Mapping of top-down methods

Top-down methods		
Estimation methodologies	**Endogenous/dependent variables**	**Application**
Basic algebraic operations with time series and panel data	Trade balance, trade volumes and values	Čejková & Zídková (2019)
	Trade mirror statistics	CASE (2015), Braml & Felbermayr (2021)
	VAT refunds	Gajewski & Joński (2022)
Econometric and statistical modelling	Trade balance, trade volumes and values	Frunza (2016)
Machine learning	Trade mirror statistics	European Commission (2017)

Source: own elaboration.

Box 2: Machine learning

Machine learning, a branch of artificial intelligence, has gained significant attention due to its ability to enable computers to learn from data and make predictions or decisions without explicit programming. It encompasses the development of algorithms and statistical models that extract patterns and insights from extensive datasets through iterative processes. Machine learning entails training computer systems to automatically learn patterns from data and make informed decisions or predictions without explicit programming. It involves the utilisation of algorithms that iteratively learn from data, enhance their performance over time, and generalise their knowledge to novel, unseen data instances. Machine learning algorithms can be divided into three groups: supervised, unsupervised and reinforcement learning. Supervised learning is conducted on labelled datasets (where the target variable is known) and is used to make predictions. Contrary to this, unsupervised learning is conducted on unlabelled datasets, with the aim of detecting hidden or underlying patterns. The last category, reinforcement learning, is neither supervised nor unsupervised and does not rely on training datasets – instead it is carried out by “rewarding” the desired behaviours and punishing or not rewarding undesired ones.

Machine learning is one of the various techniques employed in data mining to extract valuable insights from intricate data structures. Machine learning and data mining share a fundamental objective: uncovering meaningful patterns from data. Machine learning techniques, such as classification, clustering, and regression, serve as integral components of data mining. These algorithms play a crucial role in unveiling concealed relationships, identifying patterns, and making predictions.

Machine learning algorithms in econometric data mining perform diverse functions to facilitate various data analysis tasks. Some of the primary types of machine learning algorithms include classification algorithms and clustering algorithms, described in Box 4.

Bottom-up methodologies could be classified across three dimensions: detection, measurement, and extrapolation. Firstly, this approach requires the detection of “fraudulent observations” in the micro-level dataset used in the analysis (such as, for instance, data from tax returns, invoice data, or transaction networks). Tax administrations can collect such data by running audit programs. In most cases the detection of fraud is based on risk-analysis, which could use various sophisticated algorithms, however some countries also run random audit programs – these are for instance Denmark, the Netherlands, and the UK. Next, to allow for estimating the scale of the entire fraud, the instances of detected fraud need to be measured – assigned a value of forgone revenue or traded goods. Such information could either come from audit procedure or could be observed ex-post in fiscal registers (e.g., as the value of tax debt of a specific company). The last step in these bottom-up approaches is extrapolation – a "generalisation" to the entire tax base. In other words, the knowledge on detected cases and estimated individual values of MTIC fraud/forgone revenue needs to be translated to all taxpayers and the entire tax base. In the case of observations being based on a random sample, or the modellers observing the entire population, this step is straightforward (Approach 1 and 3 in Table 5). If, however, the calculations are being conducted on a biased sample of taxpayers, econometric and statistical methods need to be used to remove this bias (for instance, the two-step Heckman procedure[78]).

On top of the methods that involve the use of audit results (that do not cover the entire population of taxpayers), members of the Fiscalis Project Group identified a methodology based on individual level data that does not require audits (see Approach 3 in Table 5). This method involves two datasets available for all tax administrations – data on cross-border transactions from VIES and information gathered from VAT returns. As discussed in Section 0, the former set is expected to contain information on all cross-border transactions, whereas the latter does not contain information on transactions in fraudulent chains. In consequence, matching the value of intra-Community Acquisitions from both sets on per taxpayer basis yields a complete value of fraudulent transactions and resulting forgone revenue.

Table 5: Mapping of bottom-up methods

Bottom-up methods			
Detection of fraud	**Measurement of the scale of detected individual fraud**	**Extrapolation**	**Application**
Based on random audits	Audit results	No bias in the sample. Do not require any complicated methodology to extrapolate.	Quoted by Fiscalis (2018), likely rarely used
Based on risk-analysis system (using e.g., machine learning, TNA, traditional risk scoring)	Audit results	Econometric methods correcting for sample selection bias using, among others, risk scores, logistic regression and other	Numerous administrations, such as the Italian Revenue Agency and Hungarian National Tax and Customs Administration
Unmatched value of ICA/ICS in VIES and VAT returns		Covers all taxpayers	Used so far to limited extent but envisaged to be implemented by some members of the Fiscalis group

Source: own elaboration. Examples of application are based on Fiscalis (2018) and Fiscalis (2022).

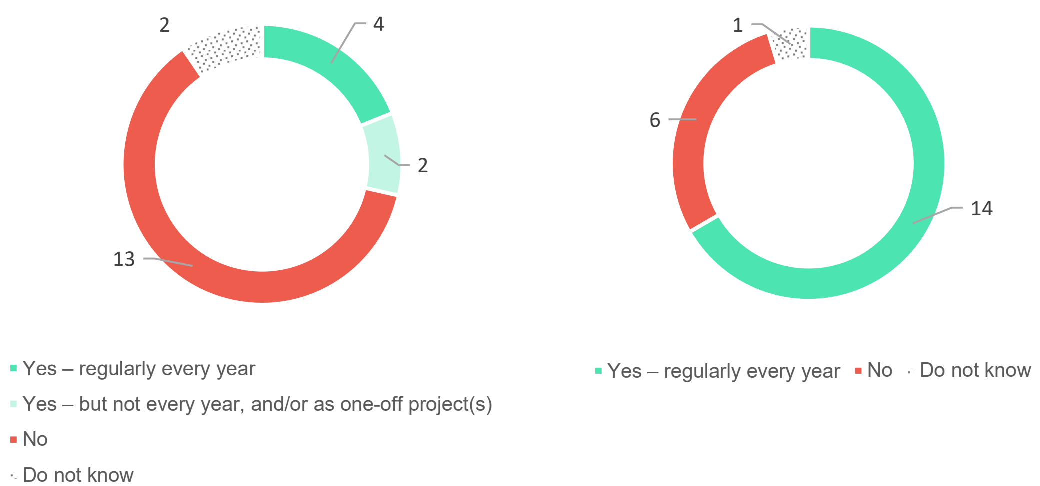
[78] The first stage of this method involves a probit analysis on a sample selection equation (the observed outcome), the results of which can then be used to predict the probability of selection for each observation. In the second a transformation of these individual probabilities predicted in the first stage is added as an additional explanatory variable in order to correct the selection bias.

II.c. Experience of Member State administrations

This chapter summarizes information gathered via the survey addressed to Member States' administrations and concerning their experience with MTIC fraud detection and MTIC fraud estimation. Answers to the remaining questionnaire questions, which pertain to the data sources used for estimation and the perceived accuracy of various estimation methods, are covered in Chapters IV and V, respectively.

As has already been mentioned earlier in this report, of the 21 Member States who responded, 14 conduct MTIC fraud detection, all on an annual basis. This suggests that most Member States take active measures to detect and combat this type of fraud. However, the responses also demonstrate that most Member States have not, at the time of conducting this survey, been in a position to estimate the actual revenue losses incurred as a result of such fraud – only six administrations have ever undertaken this task, and two of them do not do so on an annual basis (see Figure 5). These results highlight how rarely this task is undertaken and correspond with the scarcity of literature on the topic.

Figure 5: MTIC fraud measurement (left) and detection (right), as carried out by Member States

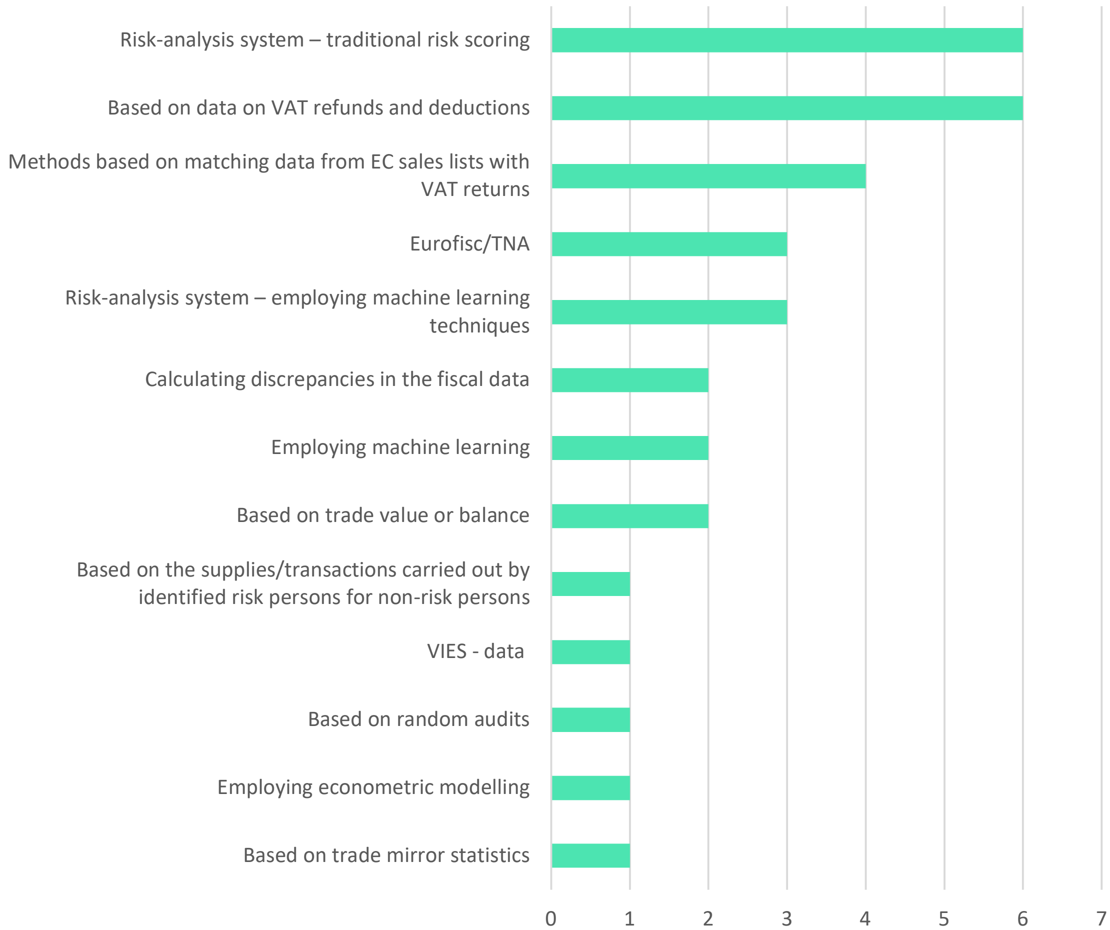


Source: Own elaboration based on survey responses.

The fraud detection and estimation methodologies offered most often in the survey responses were risk-analysis systems (bottom-up) using traditional risk scoring and methods based on data on VAT refunds and deductions (each employed by six of the 13 Member States who detect and/or measure MTIC fraud). The least common were methods based on the supplies/transactions carried out by identified risk persons for non-risk persons, methods based on random audits or trade mirror statistics, and methods employing econometric modelling (one Member State each). No administration declared using basic algebraic operations on time series and panel data for the measurement of the scale of MTIC fraud. Four administrations in total mentioned making use of EU-wide information exchange between the national tax authorities of all Member States – VIES data, under the coordination of the Eurofisc network,[79] and using the TNA (transaction network analysis) tool (see Figure 6).

[79] The network was launched in 2010, with the explicit goal of combating cross-border VAT fraud.

Figure 6: Approaches to MTIC detection/estimation used by Member States (multiple answers allowed)



Source: Own elaboration based on survey responses. Note: Based on 13 responses.

Overall, the answers gathered in the survey indicate several things. First of all, we can see that although many administrations have some kind of MTIC fraud detection program in place, most have no or very limited experience in estimating the MTIC gap. What is more, most of the methods mentioned in the survey were only employed by two or less Member States. One group of methods that stands out in particular are methods based on trade data. The methods used most often, on the other hand, rely heavily on data available to tax administrations and not necessarily accessible to the study team (e.g. audit data).

III. Assessment framework to identify a common MTIC gap estimation methodology for all EU Member States

III.a. Overall framework

This chapter discusses the evaluation framework, which was constructed having in mind the principles listed and discussed in Appendix C. The operationalisation of the framework is divided into two main steps – preselection and selection. The objective behind this two-stage process was to retain for further analysis only those methods for which data is available or can be made available during Phase II and which satisfy the minimum quality standards and minimum requirements (see Table 6 for relevant conditions). The retained methods later undergo a full semi-quantitative, semi-qualitative assessment.[80]

Keeping in mind that some criteria for assessing these approaches can be quantified and some cannot, we use a combination of Cost-Benefit Analysis and Multi-Criteria Analysis methodologies. Multi-Criteria Analysis allows for comparing alternative options along a set of pre-determined criteria. It allows one to relate the qualitative impact categories and to rank different options using more than one type of indicator, unlike Cost-Benefit Analysis. Multi-Criteria Analysis is opinion-based, as it relies on the collective opinion on the weight of different criteria. The criteria for Multi-Criteria Analysis are assigned weights, determined based on the priorities discussed and scored with the assistance of the Commission. The proposed criteria are linked to the quality and availability of information. The availability of information is viewed from two distinct perspectives: from the standpoint of Member State administrations, which do not face restrictions to accessing the data, and from the standpoint of the study team, a contractor for the European Commission. Thanks to this distinction, the assessment carried out in this study can answer two different questions, namely: (1) which approach should be chosen if it were implemented by Member State administration, and (2) which approach is optimal and feasible to be implemented by the study team, and whether it meets minimum requirements set.

As future monitoring of the MTIC gap is an important aspect of this project, on top of these criteria we assess the costs, complexity and risks involved in the continuation of this work in the future. Yet, the sub-criterion of the cost and complexity is excluded from the comparison of methodological approaches.

As the preselection stage, the assessment is restricted only to a qualitative assessment and to a narrow set of criteria, which are: completeness (across types of fraud, coverage of Member States and years), accuracy of point estimates and trend, legal and ethical risk, and continued availability of data sources in the future (see Table 6). The objective is to discard the methods that do not fully meet the objectives of this study and to group them into holistic methodological scenarios, which will undergo a full assessment.

To enable the selection of the best-suited methodological approach at the full assessment stage, decisions also need to be made with regard to the importance of specific criteria and their optimal values. These aspects are determined by two groups of interrelated parameters, which are (1) weighting and (2) scaling parameters. To scale different criteria, the study team proposed normalisation against a floor (the least desirable value, set to 0 after normalisation) and a ceiling (the most desirable value, set to 1 after normalisation).[81] This approach requires experts to determine the minimum and maximum values but, unlike other "unsupervised" methods, allows for close control of the assessment process. As an example, the methods of normalisation that are based on the actual variability of specific criteria (such as standardisation or the min-max method) could lead to an overly strong impact of the criteria that are most stable across decision options.

One of the criteria used in the preselection is related to the accuracy of the estimates generated – the confidence intervals of the figures (if the method allows) and non-statistical deviations from the estimated scale of MTIC fraud are analysed. The methods with an expected average error over 5 pp. are dropped (see Section IV.c). In addition, the study team focuses on approaches that are sufficiently complete (2/3 of the base and irregularities expected to be covered), and the comparability of time series and occurrence of structural breaks are taken into account. On top of that, ethical and legal constraints

[80] The quantitative assessment uses the criteria, their valuation and weights as described in Table 6. The qualitative assessment makes use of information gathered from the literature and the insights, perception and experiences of national administrations.
[81] See: OECD (2008).

were evaluated; the boundaries to these restrictions are less well defined and will be considered case by case. Lastly, methods based on datasets most probably out of reach for the Team were excluded from further analysis.

Preselected methodologies went through to the second stage of evaluation, in which each method was scored on the dimensions described above and using the intervals and distribution of weights. The entire approach can be summarised in tabular form. Rows in Table 6 stand for the sub-criteria used in the assessment at the preselection and/or final stage. These nine sub-criteria are grouped into three baskets: (1) accuracy, completeness, and comparability; (2) granularity; and (3) replicability. Column 3 describes whether the selected sub-criteria will be used in the preselection process and, if so, how. Column 4 provides information on how these specific criteria will be assessed and Column 5 details how they will be included in the final comparison. The latter includes the weight and scaling method. We assume that all criteria, with the exception of “Ability to link the value of the VAT fraud tax gap to specific drivers/types of fraud” and “Coverage of Member States/level of extrapolation needed to achieve coverage across EU-27”, will be normalised using a floor and a ceiling. These two remaining categories were assigned categorical scales.[82]

[82] As discussed with the European Commission, the parameters of the framework may be subject to revision at the later stage.

Table 6: Approach to evaluating preselected methodologies

(1) Criteria	(2) Sub-criteria	(3) Preselection requirement	(4) Assessment	(5) Use in a full comparison (weight and method of scaling)
Accuracy, completeness, and comparability	Accuracy of point estimates/comparability across Member States	Yes	Semi-qualitative, semi-quantitative. Errors that can be quantified (e.g., confidence intervals around parameter estimates) will be summarised using proper statistical techniques. The qualitative assessment will use insights and perceptions of national administrations.	Weight in the overall comparison: uniform distribution in the **15-25%** interval Floor: 5 pp. expected deviation on average Ceiling: expected full accuracy
	Completeness across types of MTIC fraud (directly interrelated with the above)	Yes, covers at least three-quarters of the expected scale	Qualitative – based on practitioners' expectation of the scale of fraud	Weight in the overall comparison: uniform distribution in the **10-15%** interval Floor: two-thirds covered, Ceiling: expected full coverage

(1) Criteria	(2) Sub-criteria	(3) Preselection requirement	(4) Assessment	(5) Use in a full comparison (weight and method of scaling)
	Comparability across time/accuracy of trends	Yes, <10% of observations involving a structural break in the data	Semi-qualitative, semi-quantitative. As structural breaks in the data are probably the major reason for limited accuracy over time, for each of the methods we will assess the percentage of such breaks. The qualitative assessment will use insights and perceptions of national administrations.	Weight in the overall comparison: uniform distribution in the **15-20%** interval Floor: 10% of observations classified as structural breaks Ceiling: expected full accuracy
	Coverage of Member States/level of extrapolation needed to achieve coverage across EU-27	Yes, required data available for at least 14 Member States (not necessarily accessible to the study team)	Quantitative, based on information from data sources and providers. Two alternative perspectives tested: (1) availability of data to the administrations, and (2) availability of data to the study team.	Weight in the overall comparison: uniform distribution in the **10-20%** interval Floor – no coverage Ceiling – full coverage
	Coverage of time/timeliness	No	Quantitative, based on information from data sources and providers	Weight in the overall comparison: uniform distribution in the **10-20%** interval Floor – single year from 2018-2022 covered. Ceiling – 2018-2022
Granularity	Ability to link the value of the VAT fraud tax gap to specific drivers/types of fraud	No	Qualitative	Weight in the overall comparison: uniform distribution in the 1**0-30%** interval 0: No breakdown 0.5: Possibility of breakdown by type of irregularities or type of taxpayers 1: Possibility of breakdown by type of irregularities and type of taxpayers

(1) Criteria	(2) Sub-criteria	(3) Preselection requirement	(4) Assessment	(5) Use in a full comparison (weight and method of scaling)
Replicability	Ethical and legal risks	Yes	Qualitative	N/A. Used only as a preselection criterion
	Cost and complexity of the methodology	No	Quantitative, based on the Standard Cost Model	N/A. Additional indicator. Not included in the comparison
	Continued availability of data sources	Yes	Qualitative	N/A. Used only as a preselection criterion

Source: own elaboration.

Box 3: Standard Cost Model

Within the Standard Cost Model, costs are assessed on the basis of the average cost of the required activity ("Price"), multiplied by the total number of activities performed per year ("Quantity"). The average cost per action is estimated by multiplying a tariff (based on average labour cost per hour, including prorated overheads) and the time required per action. Where appropriate, other types of costs, such as the cost of outsourcing, equipment, or supplies, should be taken into account. The quantity is calculated as the frequency of required actions multiplied by the number of entities concerned. In case of multiple relevant activities per obligation, activities need to be totalled to calculate the cost per information event, i.e., the preparation of the dataset by Member State administrations.

III.b. Accuracy criterion

The accuracy of the proposed methodological scenarios is the key assessment criterion. Accuracy is understood broadly and covers multiple aspects. It pertains to the data used and quantitative methods employed, but also to human factors.

First, the inaccuracy, or in other words error, can be decomposed into statistical and non-statistical error components. Unlike with non-statistical error, it may be possible to quantify the statistical error by analysing the distribution of estimated parameters around unknown true value. The statistical error could be decomposed into the sampling error in the dataset used and the error related to the quantitative method used.

For our purposes, it is convenient to measure the sampling error as "margin of error", equal to half of the confidence interval for the gap estimated at the 95 level of confidence (1).

$$\textit{Sampling error} = \textit{margin of error} = +/- \ 1.96*(\textit{standard error}) \quad (1)$$

One of the determinants of the sampling error is the sample size *N*: the standard error and, as a result, the sampling error decreases proportionately to the inverse square root of *N*. Other determinants of the statistical error are the factors that affect the standard deviation of the unexplained error in the statistical model, in particular the share of variation explained by exogenous variables. The presence of the sampling error, however, leads to lower accuracy when estimating year-to-year difference as uncorrelated statistical errors add up.

The non-sampling error is an error resulting from the violation of assumptions used in the model or other "non-statistical" problems with the data. Typically, such error does not diminish with sample size, and cannot be precisely quantified. The measurement error in the data could be related to non-response/non-observation or inaccurate response/observation. As an example, such an error in the trade data results from the fact that some traders are not obliged to register their transactions, or that they make accidental errors in reporting. The non-statistical error related to the use of a statistical/econometric model results from the violation of the underlying assumptions, as the endogeneity of regressors, that may be caused by omitted variables.

As the scale of the measured phenomenon is unknown, it has to be assessed indirectly, and this needs to be done from various angles. More specifically, we:

1) Judge overall accuracy based on experts' knowledge (perceived accuracy),

2) Assess sampling error for figures for which properties are known,
3) Enumerate elements of non-statistical errors and assess their potential impacts using external data.

Unfortunately, as discussed in Chapter II, the available evidence was insufficient for comparing the results of different methods or model specifications implemented in the past by the researchers, which could have yielded some insights regarding the accuracy of different approaches (could have given evidence on the lower bound of average error in earlier studies).

IV. Preselection of methodological scenarios for further assessment

This chapter describes the first stage of the assessment process, which is the construction and preselection of methodological scenarios to be retained for full-fledged assessment. The preselection process is divided into steps, which correspond to the respective sections of this chapter. In order to allow the identification of the methodological gaps, as a first step this chapter reviews data availability and the basic characteristics which make them useful or not for the estimation of the scale of MTIC fraud (Section IV.a). As a second step, this chapter identifies methodological niches, i.e. promising methods that have been neither documented in articles and reports, nor implemented by tax administrations. The combination of identified methodologies (Section II.b) and methodological gaps (Section IV.b) is then used to shortlist the analytical methods that meet the minimum criteria set by this study (Section IV.c). These methods undergo a pre-assessment (Section IV.d). Finally, the preliminary analysis of the complementarity of different methodological approaches serves the construction of comprehensive methodological scenarios consisting of different datasets and analytical tools (Section IV.e).

IV.a. Data availability analysis

One of the key considerations when choosing the optimal methodology for estimating the scale of MTIC fraud is, broadly speaking, data availability. In this section the main candidates for primary and secondary sources of data are discussed. The selection of those particular datasets is largely based on the review of previously employed methodologies and sources of information used in those approaches. The data considerations contained in this section serve to recognize the limitations of certain sources of information and explain how these datasets can be linked to observing MTIC fraud phenomena. At the end of this section the information on declared availability of certain datasets and readiness to give access to them, as declared by tax administration representatives in the circulated questionnaire, is presented.

This review is divided into two main parts: publicly available and restricted sources of data. This distinction is important – restricted datasets could not be analysed at the same level of detail as their counterparts. Information on restricted datasets is mostly indirect, based on general descriptions of their contents, documents prepared by working groups consisting of tax administration representatives, and interviews with practitioners with direct access to those datasets/systems. On the other hand, publicly available data sources allow for more in-depth evaluation, including test implementation (see Section VII).

Publicly available data

Data on intra-Community trade in goods

As pointed out in Chapter 0, one of the sources of potentially useful information for estimating the scale of MTIC fraud is the statistics on international, and specifically intra-Community, trade (such as Intrastat, UN-Comtrade, Balance of Payment statistics). The main feature of this kind of data in the context of identifying discrepancies is that transactions are reported at both ends, allowing to compile so-called mirror statistics. In theory, the information reported by two trade partners should be the same, provided that the transaction partners were compliant with the reporting obligations. Moreover, trade statistics should be relatively stable in time, as patterns of final and intermediate consumption, and of trade partnerships, could be treated as structural patterns of the economy.

Researchers have proposed various reasons for the discrepancies observed in international trade mirror statistics (Hamanaka, 2012; Republic of Slovenia Statistical Office, 2019).[83] Many of those factors are related to objective differences in reporting concepts on opposite sides of the trade flows. For example, records of imports usually include the cost of insurance and freight (CIF), while exports are usually reported on a Free-on-Board basis (FOB). It would then follow that the value of reported imports should, at least in theory, exceed the value of reported exports. This is not the case – various studies have demonstrated that, according to official statistics, the global trade runs a surplus (e.g., Braml and Felbermayr, 2021), which is an obvious impossibility. Other explanations suggested for the discrepancies found in mirror statistic include: (1) differences in rules of treatment of re-export; (2) time lag between the dispatch and the arrival; (3) distortion caused by the application of exchange rates; and (4) differences between Member States' reporting thresholds. Another factor that could contribute to a certain intertemporal shift is the deadline for reporting; Intrastat declarations should usually be submitted by the tenth working day of the month following the statistical reference month.[84] Certainly, various unintentional misclassifications, clerical errors, and non-filing are responsible for some of the discrepancies, but those should not be consistent and would therefore at most create one-off hikes in the discrepancies between reported values. On top of the aforementioned factors, some researchers suggest that deliberate misreporting and tax and customs fraud contribute significantly to the level of discrepancies in mirror statistics (Carrère and Grigoriou, 2014; Hamanaka, 2012; Braml and Felbermayr, 2021).

As observed by Braml and Felbermayr (2021), the EU contributed to 86% of the global trade surplus in 2018. Such a marked geographical concentration might suggest that this surplus is related to other EU-specific phenomena, including MTIC fraud. This hypothesis could be verified through the elimination (or suppression) of other factors suspected of contributing to mirror statistics discrepancies (listed in the paragraph above). The (1) rules of treatment of re-export are consistent within the EU; (2) the geographical proximity means that the time lag should be relatively small; and (3) the discrepancies persist even within the euro area (thus they cannot be related to the application of exchange rates). Differences in reporting thresholds could indeed be one of the factors playing a significant role in explaining discrepancies in mirror statistics, but it should be possible to adjust the statistics accordingly to account for this. This can be attempted using, for example, Structural Business Statistics and National Accounts to adjust for different industrial structure on the two opposite sides of trade flows.

---

[83] See more discussion in Chapter II.

[84] See https://tulli.fi/en/intrastat/the-due-dates-for-submitting-statistical-declarations.

The most detailed statistics of intracommunity trade are available through Intrastat, making this dataset the primary candidate for the application of methodologies based on mirror statistics discrepancies. In general, VAT-registered businesses in the EU are required to make Intrastat declarations when moving goods between EU Member States. Intrastat filing requirements apply to businesses that are making Intra Community Supply and Intra-Community Acquisition over a certain threshold each year. The threshold varies across Member States, years and sectors. The information covered by the Intrastat declaration differs for each Member State. In general, an Intrastat declaration includes the following:[85]

- Description of goods;
- Information on Member State of dispatch/arrival;
- CN8 code;
- Quantity of goods;
- Value of goods.

In the case of MTIC fraud, the company making an Intra-Community Supply has the incentive to file Intrastat declarations, provided they exceed the thresholds in their country, to maintain a cloak of legitimacy. On the other side of the transaction (Intra-Community Supply), the purchaser acts as a missing trader, and is likely to disappear without meeting its reporting obligations, including the filing of Intrastat declarations. This discrepancy between registered dispatches (i.e., Intra-Community Supply) and arrivals (i.e., Intra-Community Acquisition) on Intrastat could potentially indicate MTIC fraud. Of course, this might not always be the case – as the supplier participating in the fraud (i.e., a broker) might not make an Intrastat declaration for its Intra-Community Supply, or the missing trader might choose to record its Intra-Community Acquisition. In such situations there would be no indication of MTIC fraud in the Intrastat data.

Although this data is the most complete source of information on international trade, it is not exhaustive and there are some categories of goods and (especially) services that are not included, but are nevertheless relevant for MTIC fraud rate estimation. Assuming that categories of goods and services for which the reverse charge mechanism was introduced are those where MTIC fraud was most prevalent, there are some notable missing pieces in the Intrastat database. Among such categories, the following are not included in the Intrastat: construction work and cleaning services; immovable property; the transfer of allowances to emit greenhouse gases; gas and electricity certificates; and telecommunication services (a more detailed list of categories can be found in Section V.a). This information gap could be remedied to some extent with other sources of data, which are described in later in this section.

Data on international trade in services

As previously mentioned, statistics on trade flows based on declarations provided when moving goods across the border (such as Intrastat) are highly problematic when it comes to the coverage of services (and certain immovable goods). Although traditionally MTIC fraud is associated with the trade of goods, there is also evidence of MTIC fraud being conducted using services (such as the transfer of allowances to emit greenhouse gases). It is, thus, important to address the issue of missing data on international trade in services. It is also important to note that the compilation of statistics on international

[85] See: Deloitte – Intrastat Guide 2022.

trade in services is much more challenging than in the case of goods. WTO countries currently follow the General Agreement on Trade in Services (GATS) framework, which provides and regulates members' reporting obligations.[86] Specific information on international business relations and related transactions comes mostly from enterprises and banks.[87] This is the primary source of information for most datasets concerned with international trade in services, such as Balance of Payment statistics, Eurostat's international trade in services, and UNCTAD Statistics for International trade in services.

The level of detail in the abovementioned statistics is considerably lower than in the case of e.g., the Intrastat database – with data available only on an annual basis, and limited to around 140 categories (total, and at various level of aggregation) of services. Given these limitations, any methodology based on the detection of statistical patterns of discrepancies that might be associated with MTIC fraud is unfeasible. Nevertheless, it is still possible to construct mirror statistics and assess the differences in declared values of trade. This might be enough to approximate the scale of MTIC fraud related to the trade in services based on referential values for trade in goods produced on a more detailed dataset.

Other publicly available datasets

Other sources of publicly available data could be used as indicators of the MTIC gap or to gather variables controlling for other factors. One of the most versatile datasets in that regard is the national accounts, containing information on tax revenue, the structure of industries in each of the Member States, and trade margins, etc. The availability of National Accounts data varies by specific indicators and between Member States; the overall completeness is high, but in the event of missing data appropriate imputation procedures will be applied. Other examples of publicly available datasets that will most likely come in handy include the Eurostat series on business registrations and bankruptcies, as well as other indicators contained in its Short-term Business Statistics such as production by industry, or turnover in services, and so on. Such data sources can be used to control for general economic conditions, which might help to associate certain discrepancies with the natural business cycle rather than fraud. Eurostat's Structural Business Statistics can play a similar, supporting role – for example information on the size of companies in various industries might help to control for discrepancies arising from different reporting thresholds.

Data available for tax administrations

This subsection presents the inventory check of the data gathered by Member States and their potential availability for the study team. Overall, responses to the questionnaire had been submitted by 21 EU Member State administrations (see Appendix E).

The list of data sources available for tax administrations contains two sources that are available for all EU Member States. They are the data from EC sales lists available in VIES, and data from customs obligations.

EC sales list

Businesses are required to report their Intra-Community Supplies through an EC sales list return. In some Member States (e.g., Hungary and Spain)[88] businesses are also required to report their Intra-

[86] See: https://www.wto.org/english/tratop_e/serv_e/gatsqa_e.htm
[87] See: https://unstats.un.org/unsd/tradeserv/tfsits/DraftChapterV_29August.pdf
[88] See: https://marosavat.com/ec-sales-list-in-europe-esl/#how-often-submit-ec-sales-list

Community Acquisition. The frequency at which EC sales list is submitted depends on the thresholds established in each country.[89] EC sales lists typically include the following information:

- The customer's name.
- The customer's EU VAT number.
- The country codes.
- The value of the transactions reported.

Following the introduction of the EU VAT quick fixes in 2020, businesses are required to file their Intra-community supply in their EC sales listing in order for it to qualify for a zero VAT rate.[90]

The VAT in the Digital Age package proposes replacing EC sales list returns with a near-time digital reporting of Intra-Community supplies and acquisitions to a central VAT information exchange database (VIES) by 2028.[91]

Non-missing businesses in an MTIC transaction chain are likely to submit an EC sales list. The VAT registration number of the customer in the EC Sales return can potentially be used to trace the acquirer who is acting as the missing trader. The value of goods in the EC sales return associated with such traders could also provide an indication of the potential VAT loss incurred due to a fraudulent transaction. However, similar to the challenge with Intrastat declarations, such analysis is contingent on the submission of returns. Yet, as pointed out by experts interviewed by the study team, businesses always submit their returns, as otherwise they would not qualify for zero-rate, which would not allow them to profit from a fraudulent scheme. In addition, an important feature of the series in VIES is that it is aggregated for companies for monthly periods. In connection to this, there is no information on the type of goods traded.

Customs obligations

The Intra-Community movement of goods and services does not result in a change in their customs status, and they are therefore not subject to any custom duties.[92] However, businesses involved in the Intra-Community movement of goods may be subject to some customs obligations. These special procedural rules may exist for the movement of certain goods such as alcoholic beverages, tobacco products and energy drinks. Additionally, Customs Procedure 42 can also be involved in MTIC fraud. In light of this, customs declarations which disclose that goods and services are being moved under Customs Procedure 42 can potentially help identify fraud.

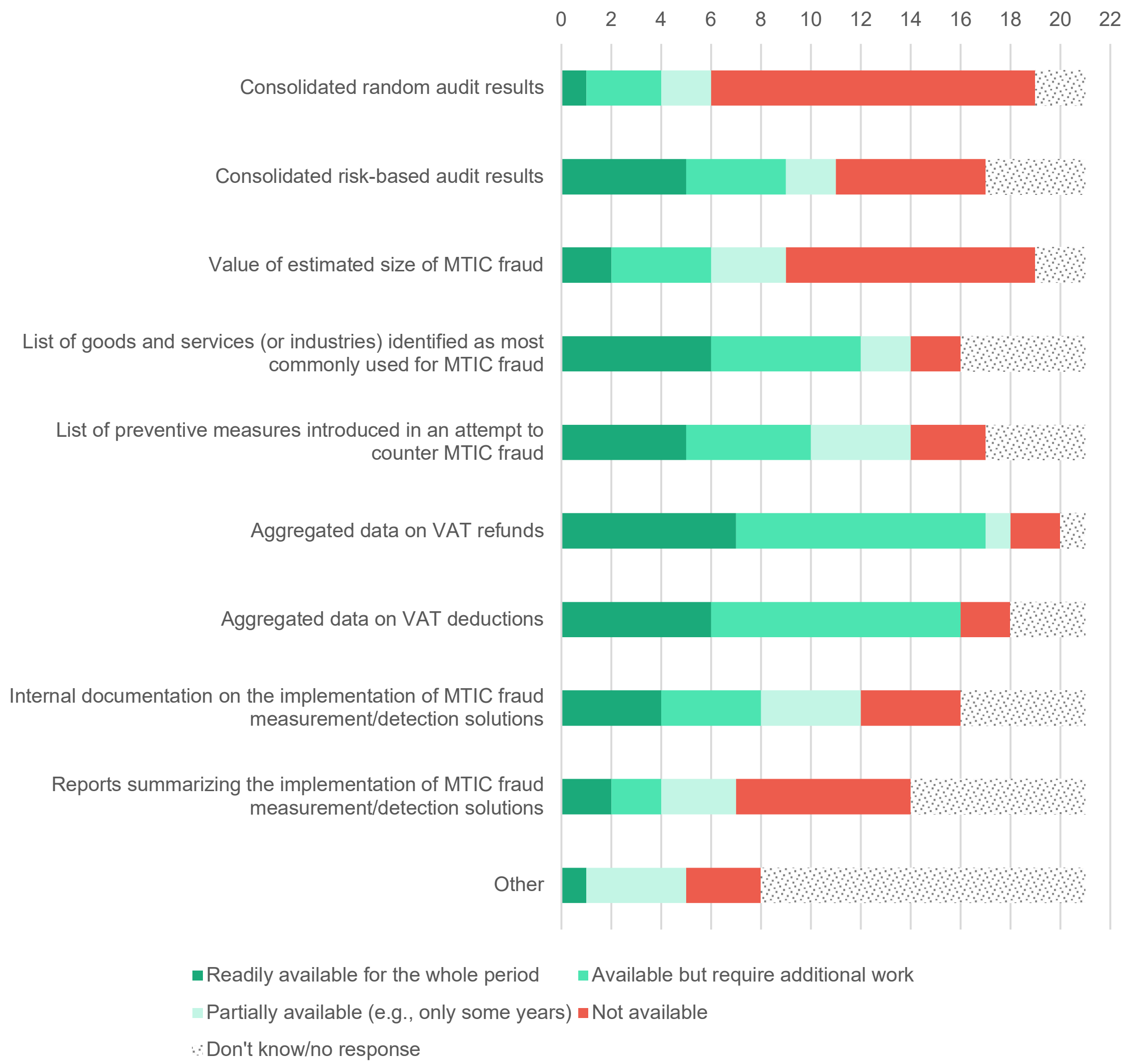
Other

As part of the survey, tax administrations were asked about the availability of certain relevant data sources. A summary of their responses is visualised below, in Figure 7. Significant challenges across the Member States were found in terms of data availability, as for each data source less than half of the respondents confirmed that the data was readily available. Consolidated random audit results and reports summarising the implementation of MTIC fraud measurement and detection solutions ranked the lowest, as the overwhelming majority of Member States said they were unavailable or only partially

[89] See: https://marosavat.com/ec-sales-list-in-europe-esl/#how-often-submit-ec-sales-list
[90] See: https://www.pwc.nl/en/insights-and-publications/tax-news/pwc-special-budget-day/2020-tax-plan-vat-quick-fixes.html
[91] See: https://www.pwc.com/gx/en/tax/newsletters/tax-policy-bulletin/assets/pwc-vat-in-the-digital-age-proposals-published-by-the-ec.pdf
[92] See: https://trade.ec.europa.eu/access-to-markets/en/content/customs-clearance-documents-and-procedures

available (did not cover the entire period under study). However, several sources were identified as available for the whole period by over half of the respondents, indicating that they might also be available for over half of the EU Member States. These were: lists of goods, services and industries identified as at risk of MTIC fraud (12 of 21) and aggregated data on VAT refunds (17 of 21) and on VAT deductions (16 of 21).

Figure 7: Summary of responses to questions on availability of relevant data sources



Source: own elaboration based on tax administration responses to questionnaire.
Note: In cases of conflicting answers, the more conservative one was chosen.

IV.b. Methodological gaps

The mapping of MTIC fraud pathways (see Chapter I) and the mapping of methodologies used to date for MTIC gap estimation (see Chapter II.b) reveal some methodological niches. In other words, they point to certain potentially promising methodological approaches and data sources that have not yet been explored or, at least, have not been documented in publicly available sources. This section discusses which data sources could prove useful and which quantitative methods may reduce some of the limitations in the earlier studies, quoted in Section IV.a.

The starting point of the analysis was **the exploration of new sources of useful direct and indirect information on the scale of fraud.** Since the issue at stake is the estimation of the scale of a largely unobservable phenomenon, the problem of finding proper indicators is of the utmost importance. The tailoring and adaptation of quantitative methods to extract this information, on the other hand, should be treated as a problem of secondary importance. With this in mind, the following data sources and indicators were selected as potentially useful for estimating the scale of MTIC fraud:

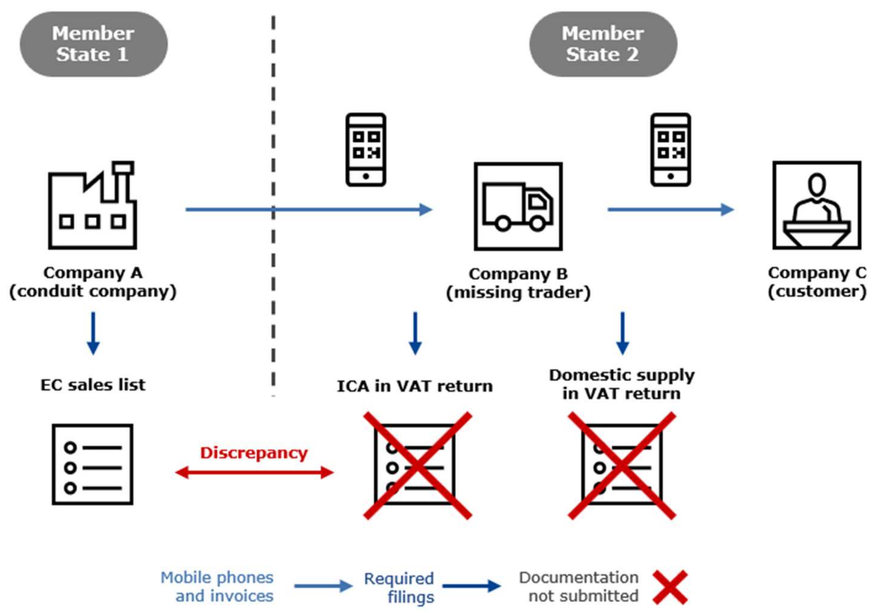
- **Administrative data from audit procedures, tax returns, unsettled liabilities, and the value of tax debt.** Audits carried out as part of the operational activities of tax administrations and proceedings, triggered in consequence of unpaid VAT liabilities, yield micro-level information on the revealed cases of fraud. Although this information does not encompass the cases that have not yet been revealed, it could be expected that administrations learn post-factum about the large fraction of MTIC fraud cases. Administrative data could be used to validate calculations using other approaches that could target more up-to-date estimates. It is important to highlight that the results of risk-based audits are often used to estimate the overall VAT compliance gap (see EC, 2022) and break it down by components, such as MTIC fraud.
- **Intra-Community trade values**. As discussed in Chapter I, fraudulent transactions in goods are expected to be partly recorded in the intra-Community trade figures reported in Intrastat (or EC sales returns). At the same time, fraudulent transactions in services are reported in the national reporting obligations system available in Eurostat's international trade in services series. The recorded part of trade in goods and services is expected to cover schemes in which the same commodities have never been moved or are moved repeatedly across borders. Since transactions in such carousel frauds are *artificial*, i.e., goods subjected to fraud do not satisfy domestic demand but are re-exported in a circular scheme, they shift values and volumes of intra-Community trade both in the missing traders' countries and the conduit companies' establishment. Shifts in trade values have already proved to be a relevant source of information in earlier studies (see e.g., CASE, 2015).
- **Intra-Community trade mirror statistics**. As discussed in Chapter I, in some circumstances (e.g., simple MTIC fraud or final movement of goods in carousel frauds), the missing trader does not have incentives to register intra-Community acquisition in the Intrastat system. As a result, discrepancies appear in trade values in mirror registers. This observation has already been used by numerous researchers (see e.g., Braml and Felbermayr, 2021). However, Intrastat only covers trade in goods, and there is no sufficiently detailed information that would enable the comparison of mirror registers of trade in services.
- **Discrepancies between trade values in VIES and VAT returns**. Since the introduction of the EU VAT quick fixes in 2020, businesses have been required to file their Intra-Community Supply in their EC sales listing in order for it to qualify for a zero VAT rate. With the discrepancies between these figures – that are thus a complete source of information on trade values – and VAT returns, the value of non-fraudulent transactions could be used to infer the value of MTIC fraud.[93] Yet, recording these discrepancies follows a complex pattern. As explained in Chapter I, the behaviour of fraudsters may vary depending on their

[93] The methodology using VIES and VAT returns was suggested by Fiscalis Project Group Members. However, to our knowledge this approach has not been documented to date.

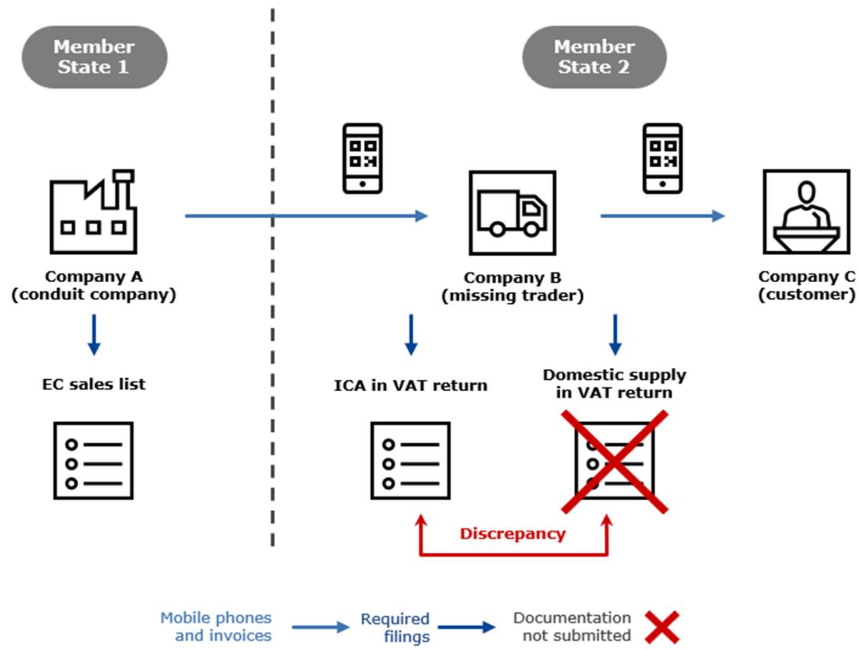
expectation of the earlier actions of tax control authorities. Most importantly, missing traders may or may not report their intra-Community acquisition (see Figure 8 (a) and Figure 8 (b), respectively), although the former is more likely. As in all cases, missing traders are expected to exclude fraudulent domestic supply from their returns, which always results in discrepancies in the register. In situation (a), a mismatch between VIES and intra-Community acquisition in VAT returns arises. In situation (b), i.e. correct reporting of intra-Community acquisition, a discrepancy between intra-Community acquisition and domestic supply filed in tax returns will arise. Importantly, in the case of the intra-Community trade in final goods, such as, for instance, mobile phones, the discrepancies in registers would be more straightforward to interpret. Meanwhile, tracking fraud in intermediate goods, such as steel products, would require controlling for the possible transformation goods and services acquired by legitimate traders. As a result, the use of discrepancies in VAT returns would require using sophisticated tools to account for evolving product codes in different value-added chains.

Figure 8: Visual representation of the discrepancies using the example of fraud involving mobile phones

(a)



(b)



Source: own elaboration.

- **VAT repayments**. Artificial transactions involved in carousel fraud involve large-scale repayments that would not be requested if fraud did not take place. As a result, MTIC fraud impacts the relation between VAT repayments, and gross and net VAT revenue (see Gajewski and Joński, 2022).
- **Overall VAT compliance gap and C-efficiency**. Forgone revenue due to MTIC fraud is an element of the overall VAT compliance gap. It is also an element of inefficiencies of VAT collection proxied by the C-efficiency ratio. As the MTIC gap is likely one of the most dynamic components of the VAT compliance gap, i.e., more volatile than the components linked to tax morale and structural patterns of the economy, the dynamics of MTIC fraud is expected to correlate closely to the VAT compliance gap (see EC, 2022). Although the VAT gap in the EU study is often used to support experts' assessment of the scale of MTIC fraud, to the study team's knowledge the unexplained fluctuations of the VAT compliance gap have never been used for estimating the MTIC gap.

One series of information initially identified as potentially useful, namely **the number and frequency of the CPC 42 used by traders**, has been discarded, as the information brought by this indicator appears to be very weakly related to the scale of fraud. This procedure appears to be an intrinsic element of MTEC fraud, rather than the schemes that do not involve imports from third countries. Moreover, as shown by the EC/CASE (2021), variation in the number of procedures used was only weakly correlated with the overall VAT compliance gap.

Although the set of indicators that could indicate, directly or indirectly, the level or dynamics of fraud is numerous, it must be kept in mind that all the above-mentioned indicators pose substantial limitations. These constraints are related to, among others, their incompleteness and complexity. As an example of the incompleteness issue, administrative data shows only the scale of revealed fraud. Hence, they pose a challenge to extrapolating forgone revenue to all taxpayers and the entire tax base. At the same time, irregularities in repayments and deductions could only indicate a fraction of MTIC fraud as they are likely a poor indicator of simple MTIC fraud cases that do not involve the circular movement of goods (as carousel fraud). Secondly, most of the listed series are *complex* and show much more than the fraud and its dynamics and, therefore, their accuracy is limited if the information is processed improperly. As an example, discrepancies in trade mirror statistics, on top of indicating the fraud level, can also be (partly) explained by differences in Intrastat registration thresholds and the distribution of companies' sizes across borders. Moving on, VAT repayments are determined not only by fraud but also by such factors as the openness of the economy to trade, structure of tax rate systems, and share of intermediate use in companies' output.

The existence of these limitations increases the need for using sophisticated quantitative methods to extract only the relevant variation of indicators. At the same time, a large fraction of research conducted up to date places emphasis on simplicity and the simple quantitative methods that were often used do not control sufficiently for the noise in the data. In addition, to our knowledge there has been no study that would combine various indicators in order to reduce the impact of their individual limitations. Neither synthetic methods using multiple indicators, nor studies combining different methods in a hybrid approach, have been proposed to date. The latter approach could theoretically be operationalised to ensure more complete coverage of the gap, allow for validation of estimates, allow for a breakdown of the MTIC gap into components, or to enable coverage of time periods for which some statistics are not available.

Controlling effectively for other factors affecting the dynamics of the indicators of fraud and combining information brought by various indicators are the two overarching objectives behind the novel approaches proposed in the following sections. These novel approaches were inspired by the broader literature dealing with similar problems, i.e., estimating unobservable phenomena based on the number of partial and inaccurate indicators. There are two well-established methodologies for dealing with such problems, known as structural equation modelling (SEM), and state-space models. These approaches share many similarities and can be shown to be special cases of one another, depending on the kinds of restrictions imposed. The main difference between the methods lies in the different modelling paradigms they exemplify (Chow et al, 2010). SEM and its specific application, Multiple Indicators, Multiple Causes Measurement (MIMIC) is described in Appendix D. The Kalman filter, which is a specific example of the state space model, well established for estimating unobservable phenomena (see also Appendix D). As shown in the following section, these methods can prove well suited for estimating the MTIC gap.

IV.c. Retained methods

Based on the mapping of MTIC fraud pathways (see Chapter I), mapping of documented methodologies (Section II.b), and analysis of the gaps in approaches used for estimating the MTIC gap thus far (Section IV.b), this section provides a typology of all quantitative methods under consideration. The methodological approaches presented differ in terms of their novelty. The set of methods includes original approaches not tested so far for this estimation, methodologies inspired by earlier research, and approaches already documented in the literature. Some of these techniques use a single indicator (of the six listed in Section IV.b), others combine information from a handful of variables. The approaches using only selected indicators form broader estimation scenarios. The following section has a discussion of the rationale behind combining these approaches and proposes concrete scenarios (see Section IV.e).

Methods based on Intra-Community trade data from Intrastat

As discussed in Chapter I, one can observe specific patterns of registering fraudulent intra-Community trade transactions in the Intrastat system and other reporting obligations. First, the conduit company has all the incentives to be compliant and register Intra-Community Supply. At the same time, a missing trader is likely not to register its Intra-Community Acquisition as it *disappears* after running the fraudulent scheme. In the case of circular schemes, registration in Intrastat will also be carried out by the broker moving the goods to the country where the conduit company is established.

Although the transaction schemes can be complicated and may involve some deviations from this pattern, like the fact of reporting the Intra-Community Acquisition, the important fact is that at some point in time the transactions are recorded in the Intrastat system and leave a trace in the statistics. Overall, two patterns in intra-Community trade data may reveal MTIC fraud and its scale. These are: (1) sudden unexplained hikes in trade values leading to relatively high trade values (compared to other Member States and types of products), (2) discrepancies in mirror statistics.

For **analysing irregularities in trade figures and for explaining the nature of these deviations**, time series econometric techniques should be envisaged (**Method #1** in Table 7). Such methods would help control for other important factors affecting trade, for instance seasonality in the demand for specific goods and services, or shifts in demand caused by the economic cycle and other events (for example the lockdown during the COVID-19 pandemic). These factors are expected to play an important role, thus accurate analysis would not be possible without the proper econometric tools.

In most published studies, **the analysis of discrepancies** in the different registers uses simple algebraic operations without controlling for the factors other than MTIC that could cause such irregularities (**Method #2** in Table 7).[94] More sophisticated approaches would need to account for other factors by grouping discrepancies into those caused by MTIC fraud and those caused by other factors. There are two potential approaches to grouping discrepancies and shifts in trade. One of them is that of **machine learning classification methods** (discrete choice econometric models and decision trees, **Method #3**) that require a training set with observed values of the endogenous variable (i.e., confirmed cases of MTIC fraud in trade data). The second group comprises **clustering techniques** (such as k-means, **Method #4**) that could be used to observe distinction in the patterns of the analysed phenomenon (discrepancies in trade) described by a series of variables (such as volatility or value of the Intrastat registration threshold).

[94] See Table 4 for reference to Table 7.

Table 7: Typology of proposed methods, which are based on trade data

Methods based on trade data		
Reference/Number	**Endogenous/dependent variables**	**Estimation methodology**
Method #1	Values of ICA and ICS	Econometric time series analysis
Method #2	Trade mirror statistics	Simple algebra
Method #3	Values of ICA and ICS	Classification machine learning techniques
Method #4	Values of ICA and ICS, trade mirror statistics	Clustering machine learning techniques

Source: own elaboration.

Under proposed Method #1, the study team would compile a dataset of trade statistics from the Intrastat system, broken down by year, product and country. The analysis would be carried out on the country rather than country-pair level due to expected fluctuations in trading partners or simply trading roots. It would probably use either the 4-digit or 6-digit CN codes level from Intrastat and all available categories of trade in services reported by Eurostat. In the less complex alternative using the 4-digit CN codes from Eurostat there would be over 4 million observations in the dataset. The value of trade in each country, month and product group would be explained econometrically depending on its past and future realisations (to account for the trend) and economic factors such as sectoral GDP growth. Large deviations from trends and patterns would be regarded as irregularities caused by MTIC fraud.

Under proposed Method #2, the study team would compile a dataset of trade statistics from the Intrastat system broken down by year, product, country-pair and Intrastat register. Similar to the analysis envisaged under Method #1, the analysis would be conducted per month and either the 4-digit or 6-digit CN codes level. In a less complex case there would be ca. 109 million observations in the dataset spanning from 2010 to the current date. The difference on the registers of goods that could have been subject to MTIC fraud would be attributed to the fraud and used to estimate the forgone VAT revenue.

Under proposed Method #3 and Method #4, the study team would also compile a large dataset of trade values and discrepancies, as well as calculate characteristics of their dynamics. This data from the Intrastat system and data on intra-Community trade in services would then be merged with economic variables that could explain volatility and discrepancies in trade data. The study team would also need to gather a training set containing dummy variable indicating periods, countries and groups of products on which MTIC fraud was observed as well as periods, countries and groups of products on which MTIC fraud did not take place: Such a **training set** could be compiled for countries and goods that introduced the reverse charge mechanism and other mechanisms eliminating MTIC fraud (see Box 2). Under Method #3, the Team would test alternative classification algorithms including discrete choice econometric models (like logit and probit specifications) and decision trees. These algorithms would be trained to classify the observations to fraudulent and non-fraudulent based on the explanatory variables describing various characteristics of trade for relevant period, type of goods, and type of product or service. The optimal method would be selected based on the so-called receiver operating characteristic (ROC) curve, depicting the interrelation between the sensitivity and specificity of the classification. Finally, the model trained on a fraction of the observations would be used to predict instances of fraud for observations not included in the training set. Under Method #4, the observation would be clustered using k-means or a similar algorithm. The k-means algorithm groups similar observations into segments, thus allowing the identification of "odd" values. Within the context of the described estimation, the "odd" values refer to the observations of fraudulent transactions. In comparison to a "supervised" classification algorithm, the prediction of which observations are fraudulent, which not, do not require a training set. Yet, as clustering algorithms are "unsupervised", to avoid detection of unwanted type of anomalies (not reflecting the instances of fraud but something else), different numbers of clusters and starting centroids would be tested. Finally, the results could be validated on the training dataset.

Box 4: Classification and clustering algorithms

Classification Algorithms

Classification algorithms are a type of supervised learning algorithm used to predict a categorical outcome based on input data. Some common types of classification algorithms include logistic regression, decision trees, k-nearest neighbours, and support vector machines.

Logistic regression or, in other words logit model, is a type of generalised linear model used for binary classification. The algorithm estimates the probability that an observation belongs to a particular class by fitting a logistic function to the data. The formula for logistic regression is:

$$p(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}}$$

where p(y=1|x) is the probability that the observation x belongs to class 1, β_0 is the intercept term, and $\beta_1,...,\beta_n$ are the coefficients for the predictor variables $x_1,...,x_n$.[95]. The probit model, on the other hand, estimates the probability of an event by fitting a cumulative distribution function (CDF) of a standard normal distribution to the binary outcome. It assumes that the log-odds of the probability follows a linear relationship with the predictor variables.

Decision trees are another type of classification algorithm that can be used for both binary and multi-class classification. The algorithm constructs a tree-like model of decisions and their possible consequences. At each internal node of the tree, a test is performed on one of the input features, and the outcome of the test determines which branch of the tree to follow. The leaf nodes of the tree represent the predicted class labels.[96] One of the most widely used decision tree algorithms is J48 and its implementation, C4.5, developed in 1994 by Ross Quinlan.

k-Nearest Neighbours is a non-parametric classification algorithm that predicts the class label of an observation based on the class labels of its k nearest neighbours in the training data. The distance between observations is typically calculated using a distance metric such as Euclidean distance.[97] The predicted class label is determined by majority vote among the k-nearest neighbours.[98]

---

[95] See Bishop, C. M. (2006).

[96] See Hastie, T., Tibshirani, R., & Friedman, J. (2009).

[97] In two-dimensional space, the Euclidean distance between two points (x1,y1) and (x2,y2) is given by the formula: $d= \sqrt{(x2-x1)^2+(y2-y1)^2}$.

[98] See Hastie, T., Tibshirani, R., & Friedman, J. (2009).

Support Vector Machines are another type of classification algorithm that can be used for both binary and multi-class classification. The algorithm constructs a hyperplane or set of hyperplanes in a high-dimensional space to separate the different classes.[99] The goal is to find the hyperplane with the largest margin between the classes.[100]

Clustering Algorithms

Clustering algorithms aim to group similar data instances based on their inherent patterns or similarities. They aid in discovering hidden structures and relationships within datasets, without the requirement of predefined classes. Clustering finds applications in diverse fields, including customer segmentation, anomaly detection, and pattern recognition. Prominent clustering algorithms, such as k-Means, DBSCAN, and Hierarchical Clustering, partition data instances into groups based on their similarity, enabling analysts to discern patterns and comprehend the underlying structure of the data.

k-means is a partition-based clustering algorithm that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The algorithm starts by randomly selecting k initial cluster centroids. It then iteratively assigns each observation to the nearest cluster centroid based on the distance between the observation and the centroid. After all observations have been assigned to a cluster, the algorithm updates the cluster centroids by calculating the mean of all observations in each cluster. This process of assigning observations to clusters and updating cluster centroids is repeated until the cluster assignments no longer change.[101]

Hierarchical clustering is another type of clustering algorithm that builds a hierarchy of clusters by either merging smaller clusters into larger ones (agglomerative clustering) or by splitting larger clusters into smaller ones (divisive clustering). In agglomerative clustering, each observation starts as its own cluster. Then, at each step of the algorithm, the two closest clusters are merged into a single cluster. This process is repeated until all observations are in a single cluster. In divisive clustering, all observations start in a single cluster. Then, at each step of the algorithm, the largest cluster is split into two smaller clusters. This process is repeated until each observation is in its own cluster.[102] The result is a dendrogram that shows the nested grouping of observations.

Density-based clustering algorithms, such as DBSCAN, group observations together based on areas of higher density in the data. The algorithm defines a cluster as a maximal set of density-connected points. Two points are density-connected if there is a chain of points between them such that each point in the chain is within a certain distance (epsilon) of its neighbours and has a minimum number of points (minPts) within its epsilon-neighbourhood. The algorithm starts by selecting an arbitrary point and retrieving all points within its epsilon-neighbourhood. If the number of points in the neighbourhood is greater than or equal to minPts, a new cluster is started with the initial point as a core point. The algorithm then iteratively adds all density-reachable points to the cluster. If the number of points in the neighbourhood is less than minPts, the point is labelled as noise. This process is repeated until all points have been processed.[103]

Method based on trade value from VIES and VAT returns

[99] A hyperplane is a decision boundary that separates the data into different classes or groups. It is an (n-1)-dimensional subspace in an n-dimensional space. For example, in a two-dimensional space, a hyperplane is a line.

[100] See Alpaydin, E. (2010).

[101] See Alpaydin, E. (2010).

[102] Ibid.

[103] See Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996).

The method described in this subsection (**Method #5**) uses sensitive information available only for Member State administrations, namely data on transaction-level intra-Community Supply (from VIES), intra-Community acquisitions and domestic sales of companies (from VAT returns). The latter information could be limited to companies acquiring goods and services from other Member States above the monetary threshold which could characterise potentially fraudulent transaction chains. Yet, the underlying transaction-level data, in addition to being sensitive, is also immense. As the policies of handling the data in various Member States and the resources available to compile it are likely to vary, the feasibility of this method could vary depending on the limitations in specific Member States. We foresee two options which could be offered to Member States when executing the calculation:

a. **matching of individual-level transaction data**. In this case, we expect that the analysis would be carried out by Member State administrations, using knowledge shared the study team. To facilitate matching, the study team would describe the rules (both in the selected statistical software/programming language and in the textual documentation). The rules would describe the details of observation matching and inter-relations/merging of data from different registers. They would describe what (key-variable, registers) and how should be matched (as exact vs. non-exact match). In order to control for an important element hindering in the analysis, that is, the transformation of goods and services in the value-added chains, the study team would provide a dictionary/set of rules for typical transformations for different product and services categories. Finally, after the implementation of the rules prepared by the study team, unmatched transactions in VIES and VAT returns, along with intra-Community and domestic transactions in VAT returns, would be classified as fraudulent and used for the estimation of forgone VAT.

b. **matching of tabulated data by the study team**. In case the matching of the most granular transaction-level data could n be performed by administrations, aggregated data would be requested. The level of aggregation would depend on the restrictions of Member States' administrations on sharing the data. In case there are no restrictions preventing such a break down, similarly to Method #1, 4-digit or 6-digit CN level-data and a monthly or quarterly series would be used. Unlike in case a., matching would be performed on observations denoting specific time-periods and groups of products and services, using identical rules. Similarly to "a.", unmatched values of cross-border transactions in VIES and VAT returns, along with intra-Community and domestic transactions in VAT returns would be classified as fraudulent and used for the estimation of forgone VAT.

Methods based on VAT repayments

Since an intrinsic element of the carousel-type fraud involving circular movement of goods cross-border is that of zero-rated Intra-Community Supply, the fraud could be accompanied by change in the relative share of VAT refunds (compared to net or gross VAT revenue). This would hold in situations when brokers generate nil or low output VAT.

Being mindful of method's limitations, based on the work of Gajewski and Joński (2022), we short-list the analysis of trends in the series of repayments as a method to calculate the scale of MTIC fraud (**Method #6**). Under this method, the study team would analyse deviations in the ratio of VAT repayments and net VAT revenue. The study team would calibrate the value of the ratio that should be observed if no MTIC fraud was committed, separately for each Member State. This would need to use the ratio of repayments and revenue from the period when MTIC was believed to be less prevalent than in the following years.

In addition to this simple approach, the study team proposes to short-list for further evaluation the method that would hinge on econometric modelling of the ratio of VAT repayments and VAT revenue (**Method #7**). Under this approach the endogenous variable (the ratio of refunds and revenue) would be modelled using econometric panel data specification. As an exogenous variable, the model would use indicators related to the VAT rates systems, economic cycle and structural economic characteristics of Member States. Since country-specific ratios of VAT refunds and revenue could be prone to missing variables due to unobservable factors such as fear of audits being triggered after requesting a refund, fixed-effects specification of the model would be likely. In consequence of using such a specification, the method would reveal the variation of the scale of the fraud in time rather that than the scale itself. However, changes in the scale of fraud between periods could be indirectly used to infer a lower bound estimate of its magnitude.

It is important to note that neither of the above methods, in contrast to the methods based on trade figures, would allow for the MTIC gap to be broken down by groups of products. The analysis would also likely show only the scale of circular schemes as the simplest MTIC fraud is not expected to distort the value of VAT refunds.

Method based on changes in the overall gap

According to the various estimates quoted in Chapter II, the MTIC gap could have even accounted for more than 50% of the overall VAT compliance gap. The MTIC gap is likely one of the most dynamic components of the VAT compliance gap, more volatile than the components linked to tax morale and structural patterns of the economy.

EC/CASE (2022) shows that the employed panel data specifications explaining the dynamics of the VAT compliance gap were able to account for only up to 40% of the variation. Apart from inevitable measurement errors, a large chunk of this unexplained variation could be caused by MTIC fraud. Thus, the method based on unexplained dynamics of the overall VAT compliance gap could be used for estimating shifts in the scale of MTIC fraud (**Method #8**). This method would use the econometric specifications employed in the EC/CASE (2022) study and attribute unexplained variation of the gap to MTIC fraud.

The method analysing unexplained variation of the overall gap is likely prone to significant error. However, it could prove useful for the validation of results obtained using other indicators. The inclusion of such a method was also suggested by one of the external reviewers.

Method based on risk-based audits

Method #9 assumes the use of these results and individual level data from tax returns to model bias in the risk-based sample of taxpayers. The choice of the method used (such as the Heckman model or propensity score matching) would depend on the method used by respective administrations in selecting taxpayers for audits. It would also depend on the availability of individual level information on risk scores. After fitting the model predicting the probability of companies being audited and their non-compliance ratio, it would be used to simulate the value of irregularities for all taxpayers. It should be noted that the quality of data and efficiency of audits behind the assessment could differ in time and across Member

States. Depending on the information shared in audit results, it is likely that the method could enable the MTIC gap to be broken down into components.

It is important to note that, in contrast to Method #5 based on VIES and VAT returns, Method #9 assumes that the processing of individual-level data would be carried out by the study team. In the case of this method, such an approach would not be practical, as defining the specific approach to economic modelling requires a detailed analysis of the granular data by the study team.

Multiple indicator methods

As discussed in the preceding chapter, there are methodologies that allow modelling latent or unobservable variables with one synthetic analytical approach. Two potential approaches were shortlisted: structural equation modelling (**Method #10**) and the Kalman filter (**Method #11**) (see Section IV.b).

Under these approaches, the study team would build and test different specifications of econometrics observation/indicator equations using variation in the available indicators of MTIC fraud (likely using quarterly series) including the following indicators: (1) the value of trade in risky goods and services; (2) discrepancies between mirror statistics; (3) the ratio of VAT repayments to revenue, and (4) the overall VAT compliance gap. Similar to the EC/CASE (2022) study, the modelling would use a broad set of factors, other than MTIC fraud, suspected of partially explaining the variation in these variables. These exogenous variables would include economic parameters (e.g., GDP growth), structural factors (e.g., openness for the economy to trade), policies in place (e.g., breadth of the application of Domestic Reverse Charge Mechanism) and technical factors (e.g., the relative value of Intrastat registration thresholds).

The summary of methodologies and their basic characteristics are summarised in Table 8.

Table 8: Summary of methodologies

Method	Name	Group	Main variable and characteristics analysed	Major sources of information	Quantitative method
1	Econometric analysis of trade fluctuations	Methods based on Intra-Community trade data from Intrastat	I-C trade values and their volatility	Eurostat	Econometric time series analysis
2	Simple analysis of mirror statistics		I-C trade mirror statistics	Eurostat	Simple algebra
3	Classification algorithms to analyse trade statistics		I-C trade values and their volatility, trade mirror statistics	Eurostat	Discrete choices econometric models (as logit and probit specifications), decision trees (machine learning technique)
4	Clustering algorithms to analyse trade statistics		I-C trade values and their volatility, trade mirror statistics	Eurostat	Clustering data mining/machine learning techniques as K-means

Method	Name	Group	Main variable and characteristics analysed	Major sources of information	Quantitative method
5	**Scrutiny of discrepancies in VIES and VAT returns**	Method based on trade value from VIES and VAT returns	I-C trade values, VAT liability on ICA	Administrative data, VIES	Simple algebra
6	**Simple analysis of irregularities in VAT repayments dynamics**	Methods based on VAT repayments	VAT repayments	Administrative data	Simple algebra
7	**Econometric analysis of irregularities in VAT repayments dynamics**		VAT repayments, tax systems and economic characteristics	Administrative data, Eurostat, DG Taxud's database	Econometric panel data techniques
8	**Unexplained variation in the overall VAT compliance gap**	Method based on changes in the overall gap	Series of the VAT compliance gap	*VAT gap in the EU study*, Eurostat and other publicly available sources	Econometric panel data techniques
9	**Risk-based audit methods**	Method based on risk-based audits	Audit results, individual-level data from tax returns	Administrative data	Econometric and statistical tools
10	**Structural equation modelling**	Multiple indicator methods	Trade volumes, their volatility and mirror statistics, refunds, deduction, overall VAT gap	Eurostat, VAT gap in the EU, administrative data	Structural equation modelling
11	**Kalman filter**		Trade volumes, their volatility and mirror statistics, refunds, deduction, overall VAT gap	Eurostat, VAT gap in the EU, administrative data	Kalman filter

Source: own elaboration.

IV.d. Pre-assessment

This section provides an early assessment of the methods enumerated in the preceding section. Its main objective is to limit the full assessment only to those methods which meet the basic objectives of this study and to gather sufficient knowledge to group the methods into holistic methodological scenarios consisting of complementary tools. This pre-assessment is carried out using a narrow set of sub-criteria, as set out by Table 9.

- **Completeness.** The methods based on Eurostat series and overall VAT compliance gap (Methods #1, #2, #3, #4, #8 and partially #10 and #11, Table 8), are based on datasets that are available for all EU Member States for the period going beyond the 2017-2021 interval. Availability of VAT repayment statistics will also likely enable the minimum threshold to be met (see Section IV.a for an analysis of availability for and ability to share the data by the administrations). This is necessary to operationalise Method #6 and #7 and provide an additional indicator for Method #10 and #11. The data from VIES and VAT returns (Method #5) are available to all interviewed Member State administrations, but only as of 2020 onwards. Risk-based audit results (Method #9) are expected to be available for the majority of Member State administration, as suggested by 11 out of 17 respondents confirming their availability.

- **Accuracy of point estimates and trends.** There is a risk of single macro-indicators (Methods #1 to #2, #6 to #8) being insufficiently accurate to provide reliable estimates of the MTIC gap. There is also a risk of declining accuracy over time. Specifically, due to the evolution of schemes where both broker and buffer companies are in a position to not claim refunds in order to not attract an audit, the methods based on VAT repayments (Method #6 and #7) would likely be somewhat inaccurate for the recent periods. Moreover, using a single indicator would also cause a black-box effect, i.e., uncertainty about the accuracy and driver of the estimates. This also applies to Method #11 and Method #12, which use multiple aggregate indicators and also do not allow one to track sources of fraud. Although the statistical error involved in calculations using macro indicators would likely be modest, as shown by the literature and earlier studies, there would be uncertainty related to the understating of fraudulent behaviour, the human error involved in handling massive data sets, and the quality of the data. Since the actual scale of the fraud in its entirety is unknown, any of the short-listed approaches using macro-indicators would need to be validated using other series and training sets (i.e., observations of changes in the value of indicators before and after the implementation of anti-fraud measures). Methods #3 and #4, which are intended to use multiple characteristics of the intra-Community trade rather than aggregate indicators, are expected to reach substantially higher accuracy and allow, to some extent, explaining the shifts in the magnitude of fraud (by pinpointing groups of products and trading country-pairs). Bottom-up methods based on administrative data (Method #5 and Method #9) appear to be characterised by the highest precision. This is owing to the fact that they use actual measurements of the scale of individual fraud which, in cases where sufficient information is available, can be extrapolated to the entire tax base and all taxpayers. Yet, the accuracy of the methods based on audit results would be contingent on the effectiveness of tax administrations, which cannot be precisely assessed by the study.

- **Ethical risks.** Using a single indicator for estimating the MTIC gap may raise some ethical concerns related to the impact of the study on fraudsters' behaviour. This problem appears to be most relevant for Methods #1 to #4 and Method #6, based on single baseline indicators whose values directly depend on fraudsters' behaviour. As explained in Section I.a, companies in fraudulent chains other than missing traders face incentives to register transactions in Intrastat. If the methodology of estimating the MTIC gap is based solely on this pattern, fraudsters may change their behaviour. This would create problems for assessing MTIC fraud by Member State administrations in the future. For this reason, the approaches using more than a single characteristic of trade data or a handful of indicators minimise ethical risks. The use of VIES and VAT returns as a source of information on MTIC fraud may also somewhat impact the behaviour of fraudsters. More specifically, it may affect fraudsters' willingness to include fraudulent intra-Community acquisitions in their VAT returns. Yet, as the method also analyses the discrepancies between VAT returns (intra-Community acquisitions and domestic supplies), i.e., the method looks holistically at reporting obligations, the risk of affecting fraudsters' behaviour is low.

- **Legal risks.** There is a substantial legal risk involved in MTIC gap estimation concerning the handling of individual-level administrative data. Even if individual-level audit results and information from tax returns do not contain companies' names and VAT numbers, they provide very detailed information to the extent that they cannot be regarded as anonymised. The use of such sensitive information obtained by the study team creates legal consequences if the information is leaked (e.g., due to data mishandling). For this reason, many administrations are not allowed to share information with any other entities. This risk will be evaluated further and discussed with the Commission. This may result in the decision that the methods using individual administrative data need to be discarded. Importantly, only one Method #9 (Risk-based audit methods) involves handling individual-level taxpayer data. By design, Method #5, which is also based on administrative data, involves Member State administrations processing sensitive granular information about taxpayers and sharing either somewhat aggregated figures or the final results of their own data processing.

- **Continued availability of data sources**. No risk was detected of discontinuation in Eurostat data sources. However, there is some risk that the VAT compliance gap estimates source from the *VAT gap in the EU* study might become less accurate or be discontinued. The reason for this is the discontinuation of the Own Resource Submissions (ORS), which were the primary source of information for estimating the parameters of the VAT compliance gap model for this and earlier studies. In consequence, to minimise potential problems related to discontinuation of this information, the VAT compliance gap series could only be used as a secondary method or one of many indicators in a synthetic approach.

Table 9: Summary of early assessment

Method	Name	Completeness	Accuracy of point estimates and trends	Lagal and ethical risk	Continued availability of data sources in the future
1	**Econometric analysis of trade fluctuations**	Complete	Moderate	Low risk of affecting behaviour of fraudsters	No
2	**Simple analysis of mirror statistics**	Complete	Moderate	Low risk of affecting behaviour of fraudsters	No
3	**Classification algorithms to analyse trade statistics**	Complete	Moderate to high	Low risk of affecting behaviour of fraudsters	No
4	**Clustering algorithms to analyse trade statistics**	Complete	Moderate to high	Low risk of affecting behaviour of fraudsters	No
5	**Scrutiny of discrepancies in VIES and VAT returns**	Partially complete (from 2020 onwards)	High	No	No
6	**Simple analysis of irregularities in VAT repayments dynamics**	Partially complete (not for all Member States data is available)	Low	Low risk to affecting behaviour of fraudsters	No
7	**Econometric analysis of irregularities in VAT repayments dynamics**	Partially complete (not for all Member States data is available)	Low to moderate	Low risk to affecting behaviour of fraudsters	No

Method	Name	Completeness	Accuracy of point estimates and trends	Lagal and ethical risk	Continued availability of data sources in the future
8	Unexplained variation in the overall VAT compliance gap	Partially complete (not for all Member States data is available	Low	No	No
9	Risk-based audit method	Partially complete (not for all Member States data is available)	High	Risk to handling individual level taxpayer data	No
10	Structural equation modelling	Complete (but not for all indicators)	Moderate	No	No
11	Kalman filter	Complete (but not for all indicators)	Moderate	No	No

Source: own elaboration.

IV.e. Methodological scenarios

As discussed in the preceding sections, we expect that Method #1 to Method #4 and Method #6 to Method #8 are somewhat partial or there are risks to the precision of relevant estimates. Moreover, the use of trade data as a sole source of information creates ethical risks. For this reason, the approaches must be used in a hybrid manner unless Structural equation modelling (Scenario #6 in Table 10 or the Kalman filter (Scenario #7) is implemented. For hybrid approaches, we assume the scenario consisting of simple approaches (Combination #1) and two scenarios assuming the use of more sophisticated tools (Scenario #2 and Scenario #3). The latter ones differ by the type of algorithm used for trade data analysis (classification vs. clustering). Although it is still uncertain whether data could be received and the legal risks involved in handling the data could be minimised, the scenario using risk-based audit methods (Scenario #5) has been maintained for further assessment.

Under the hybrid scenarios, Scenarios #1-#3, the interaction between the methods has a twofold character. As the methods included under different scenarios differ in their scope, they are somewhat complementary. Specifically, the analysis of trade fluctuations and VAT repayments is expected to cover repetitive-carousel frauds, where goods are not consumed after every crossing of the border. At the same time, mirror statistics and the overall variation of the gap are expected to cover all the schemes. The comparison of the results obtained with different methods could allow to learn the prevalence of basic schemes. In addition, the estimates obtained with the use of different methods will also be used for cross-validation. Such a role is primarily envisaged for the analysis of unexplained variation in the overall VAT compliance gap and the analysis of repayment dynamics.

Table 10: Summary of scenarios

Combination	Method a	Method b	Method c	Method d	Justification
1	Econometric analysis of trade fluctuations	Simple analysis of mirror statistics	Simple analysis of irregularities in VAT repayments dynamics	Unexplained variation in the overall VAT compliance gap	Econometric analysis of trade fluctuations and simple analysis of mirror statistics used as a baseline complementing each other in estimating the scale of fraud using simpler and more complicated schemes. Other

Combination	Method a	Method b	Method c	Method d	Justification
					methods would be used for the validation of baseline results.
2	Classification algorithms to analyse trade statistics	Econometric analysis of irregularities in VAT repayments dynamics	Unexplained variation in the overall VAT compliance gap	X	Classification algorithms to analyse trade statistics used as a baseline tool. Other methods would be used for the validation of baseline results.
3	Clustering algorithms to analyse trade statistics	Econometric analysis of irregularities in VAT repayments dynamics	Unexplained variation in the overall VAT compliance gap	X	Clustering algorithms to analyse trade statistics used as a baseline tool. Other methods would be used for the validation of baseline results.
4	Scrutiny of discrepancies in VIES and VAT returns	X	X	X	Scrutiny of discrepancies in VIES and VAT returns as a baseline tool.
5	Risk-based audit methods	X	X	X	Based solely on administrative data that are expected to allow for more accurate estimation than single macro-level statistics.
6	Structural equation modelling	X	X	X	Structural equation modelling for analysing synthetically all available indicators of fraud.
7	Kalman filter	X	X	X	Kalman for analysing synthetically all available indicators of fraud.

Source: own elaboration.

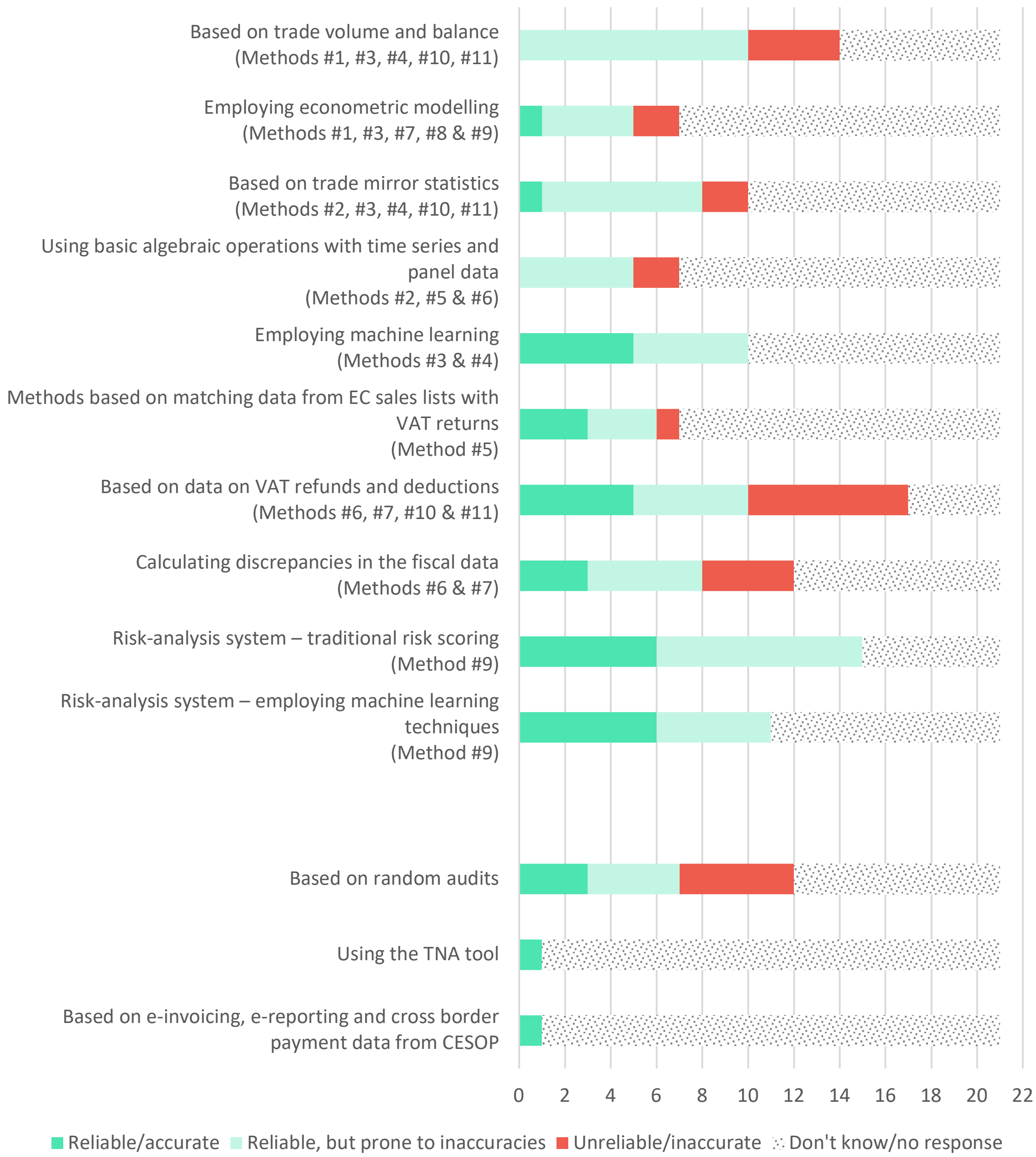
V. Full assessment

V.a. Accuracy

Perceived accuracy

The respondents to the questionnaire for tax and statistical authorities were asked to judge the accuracy of different methods used for MTIC fraud detection and estimation (see Figure 9). Based on the responses we can tentatively conclude that risk analysis systems (both using traditional risk scoring and machine learning techniques) are deemed the most reliable by the tax administrations in the surveyed Member States. Alongside methods employing machine learning, these were the only other methods not seen as unreliable by any of the respondents. However, the results of the survey come with several limitations which must be kept in mind and necessitate exercising caution when drawing conclusions. First of all, it must be kept in mind that for each method a significant share of the respondents was unable to provide an evaluation (for instance, only six respondents shared their views on the accuracy of econometric modelling methods). It is also worth noting that the views on the reliability of each method often diverged, making it difficult to determine which methods rank higher. Finally, the assessment was potentially subject to biases – for instance, we observed a degree of correlation between the most "popular" methods and the methods which were deemed the most trustworthy by a given respondent (e.g., risk analysis system methods using traditional scoring). While there could be several explanations for this connection, there is a possibility that respondents perceived the methods they themselves employed as more accurate.

Figure 9: Views on the accuracy of different methodologies used for MTIC fraud measurement and detection



Source: Own elaboration based on survey responses.
Note: Based on nineteen reposes.

Sampling, specification and estimation error

In the series that are used across various proposed scenarios, sampling error is only present in **national accounts** used to estimate the VAT compliance gap and providing indicators for multi-indicator methods.

To the extent that the national accounts total figures are based on underlying household surveys, they would be subject to sampling error, although it is rarely mentioned in this context. The sampling error for the income figures in national accounts is expected to be in the range of +/- 1-2 pp. As pointed out by the accuracy assessment of national accounts by the UK's Office of National Statistics, survey data contribute to approximately 60% of GDP calculations, whereas 40% is based on sources covering the entire tax base. As the margin of error for the Family Expenditure Survey is around 2 pp., the sampling error on national accounts (GDP) is around +/- 1.2 pp. This is in line with the observations presented in the Draft Final Report for the Personal Income Tax (including Social Security Contributions) Gap study – the average sampling error in household budget surveys is +/- 2.4 pp. Assuming some variation in the sample sizes of HBSs and different reliance of GDP calculations based on survey data, it is plausible to assume the sampling error for the national account figures is in the range of +/- 1-2 pp.

In a broader sense, the statistical error could also take the form of specification and estimation error inherent to all statistical and econometric models. Specification error arises when the model is not correctly formulated or does not capture the true relationship between the variables. Estimation errors occur when the model parameters are not accurately estimated (also to sampling error), or the assumptions of the estimation method are not met. Unfortunately, these errors cannot be estimated before the actual estimation of the models. In addition, lack of earlier estimates and published papers applying the approaches proposed under the scenarios makes it not possible to accurately assess such an error present in nearly all methodological scenarios except for scenario #4.

Other inaccuracies

For nearly all scenarios (except for Scenario #4 and #5), the analytical problem that is solved can be characterised, broadly speaking, as a problem of anomaly (or outlier) detection in time series data. This can be best seen (also because of the fact that the data is publicly available) with methods based on trade data. Even the simplest of the considered methods, where all discrepancies in trade mirror statistics are treated as the product of the fraud, can be characterised as anomaly detection (we just treat any difference between values reported on both sides of the transaction as an anomaly so there is no need for application of any sophisticated method of detection). As described in the next part of this section, this is most likely too simplistic an approach, which will not account for some of the clear shortcomings of the Intrastat data (or any other datasets where two reported values are merged and compared). A more restrictive kind of approach is most probably needed for correctly identifying the anomalies.

The optimal method of anomaly detection in time series data depends on many factors. First of all, the researchers have to make certain assumptions on the type of anomaly that is expected (or suspected). In "Anomaly Detection in Time Series" by Borges at al. (2021), a review of basic types of methodology, the authors distinguish three types of anomaly in time-series data: point (where individual observations diverge from the general pattern); contextual (where observation is categorised as anomaly in the context of neighbouring values or time of occurrence) and collective (where anomaly is characterised as unusual group of observations rather than a single point). Again, using the example of discrepancies in trade mirror statistics, it might be hard to identify contextual or collective discrepancies – as the trade data behaves too randomly and does not follow strict patterns (the seasonal component is present but not in all categories of traded goods).

Borges at al. (2021) distinguish three basic types of anomaly detection approaches to time series data: statistical-based (such as Autoregressive Model), clustering-based approaches (such as k-means algorithm) and matrix profile techniques. Statistical-based methods work best at detecting contextual anomalies, and (at least in the case of analysing trade mirror statistics) are probably not applicable. Clustering techniques are most likely the most promising as they offer an unsupervised and elastic method of anomaly detection. Matrix profile techniques represent quite a novel approach, first introduced in 2016, although the applicability of it to problems such as fraud detection has not been investigated.

Traditionally, in the field of machine learning, the algorithm is trained on a specially prepared dataset containing both features and classifiers. Such a dataset is then split into two parts – a training dataset (then sometimes split into training and validating datasets) and a test dataset. The algorithm is trained to estimate the parameters for given features based solely on the information contained in the training dataset. Fitted parameters are then evaluated using test data which was held-out from the algorithm at the training stage. A simple way of evaluating classifier prediction quality is the confusion matrix, which summarises true positives/false positives and true negatives/false negatives. If the performance of such a model is satisfactory it can be taken out to the “real world” and used to predict unclassified data. In many practical applications where machine learning algorithms are used to identify fraudulent behaviour, the training (and test) dataset can be prepared fairly easily because the classifiers can be learned post factum. For example, when a bank customer falls victim to a fraudulent transaction, in most cases they will notify the bank – which allows a particular transaction to be flagged as fraud.

In the context of MTIC gap estimation (for example based on but not limited to trade mirror statistics) this poses a significant obstacle – there is no readily available information on the occurrence of fraud, its value, length, or product categories that were the object of that fraud. Still, it might be possible to produce such a dataset if some Tax Administrations (or bodies such as EUROFISC) collect historical data on detected fraud and would be willing to share that data.

In the European Commission (2017) report, the k-means algorithm was employed for MTIC gap estimation. K-means clustering is one of the most explored techniques for grouping data and is relatively simple to implement. However, there are certain downsides to that approach: the algorithm needs to be initialised with an arbitrary number of k centroids (which will result in finding that exact number of distinctive groups, regardless of their relative difference), and the path of the algorithm depends on the initial starting points (which are chosen randomly), meaning that the algorithm might converge on a local but not necessarily global optimum. There are ways to mitigate those problems – multiple runs of the algorithm will limit the possibility of identification of a local optimum, and additional restrictions on distance between centroids should ensure that only clusters significantly different to the rest of the data points are treated as anomalies.

A big advantage of clustering techniques (k-means or others) is that they do not explicitly follow the abovementioned requirement. The clustering is an unsupervised technique of classification meaning that the dataset does not have to contain classified (or labelled) observations – the observations are essentially split into groups simply based on their relative similarity (for more see Box 3). In general, the evaluation of clustering performance can be conducted using extrinsic or intrinsic measures. Unfortunately, the extrinsic measures are simply based on comparison with classifiers, which again brings back the problem of insufficient information on when the MTIC gap actually occurred. Intrinsic measures (as the name suggests) do not rely on any other information than that used for clustering (features). The measures such as Silhouette Coefficient can help to identify consistency and separation of clusters and verify whether product clustering is not an artificial operation of splitting the whole into parts. Although useful for verifying if the clusters are sufficiently different from one another, such measures will not help to validate if the clustering is indeed related to the MTIC fraud or some other type of anomaly.

Without sufficient information on past occurrences of MTIC fraud, a large portion of the methods hinge on certain unverifiable assumptions. In the case of the trade mirror statistics example this is that fraud occurs when the values reported by two trade partners diverge (plus additional assumptions on the characteristics of those discrepancies (scale, length, pattern)). In such a scenario there are no formal means of evaluation of the produced estimates – there is simply nothing to verify the estimates against. Still, there are some contextual means of evaluating the results such as comparing the values produced on a subset containing only categories of products under a reverse charge mechanism (or the same categories before and after its introduction), comparing goods that are deemed risky and those that almost certainly are not the object of the fraud, and verifying discrepancies against, for example, publicly available information on the actions of authorities aimed at MTIC fraud (information on arrests, conducted investigations, etc.). Overall, the lack of a strict reference point for the machine learning base methods means that this approach should be considered experimental.

The limited availability of data allowed us to assess only certain methods – and the analysis below covers Intrastat data, the methods based on the value of refunds, and the method using unexplained variation of the VAT compliance gap. The analysis based on that dataset should be useful for discussing (at least some) general concepts and issues that can occur when using alternative methods.

Accuracy issues specific to selected datasets

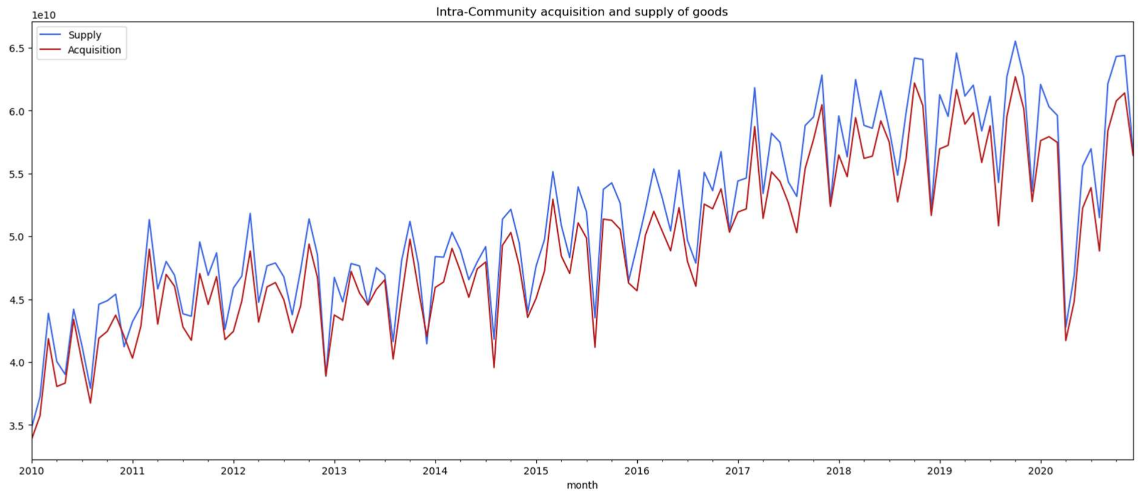
Intrastat data

Given that trade data are publicly available and often employed for this purpose, their limitations have been thoroughly addressed in existing literature. For instance, Loureiro et al. (2019) noted that the large volume of Intrastat transactions and the error-prone reporting process lead to outliers and noise that make identifying genuine anomalous data points difficult. On the other hand, Polanec et al. (2022) caution that moments estimated from the Intrastat sample, such as the mean and variance, are likely biased due to the truncation of distributions from minimum reporting thresholds. However, many of these limitations can be overcome with the application of various methods. Some of the methods used up to date to address them include hierarchical clustering methods to isolate outliers (Loureiro et al., 2019), supplementary data integration (Kľúčik, 2012; Hudec, 2013; Polanec et al., 2022), and advanced modelling like genetic programming for trade estimates to estimate missing values by adding patterns (Hudec et al., 2013).

For the purpose of a preliminary investigation of the Intrastat International Trade dataset, the study team went beyond the literature review and investigated patterns in the actual data. The analysis is limited to seven Member States of different size and from different geographical areas (Germany, France, Poland, the Netherlands, Hungary, Latvia and Malta). Although the International Trade dataset allows one to operate on an 8-digit level of the Combined Nomenclature classification, for the purpose of this exercise the analysis was limited to the 4-digit level, which is still detailed and covers 1 227 different categories of goods and services. This subset of the Intrastat database consisted of 6 506 721 observations (using rough estimation this should mean around 109 million observations for the full set of countries). In terms of general completeness of the data, 85% of the sets for each reporter, partner and CN code are present in the dataset. The remainder of the missing information can be attributed to the fact that not every product is traded between all country pairs (this is especially true for smaller Member States). The total value of export and import between country pairs (limited to the list mentioned above) is plotted in Figure 10. Here the surplus of Intra-Community Supply (I) over Intra-Community Acquisition is clearly visible. Because values are reported as totals, the discrepancy cannot be due to the misclassification of product codes. In total, the difference between the value of Intra-Community

Supply and Intra-Community Acquisition over the analysed period amounted to less than EUR 300 billion or 4.1% of the value of Intra-Community Supply.

Figure 10: Total declared export (ICS) and import (ICA) between selected Member States (Germany, France, Poland, the Netherlands, Hungary, Latvia and Malta), 2010-2020



Source: Own elaboration based on Intrastat EU trade since 1988 by HS2-4-6 and the CN8 dataset

In order to quickly check whether there is any sign of a connection between the size of the discrepancies and the categories that might be more prone to MTIC fraud, we compiled the same statistics but limited to just the product categories covered by the Reverse Charge Mechanism (RCM). At this stage those categories are considered to be at the highest risk of being subject to MTIC fraud. In the European Union, the Reverse Charge Mechanism is a mechanism which shifts the responsibility for paying VAT from the supplier to the recipient of goods or services, and is a measure used primarily to prevent tax evasion (and fraud). The Reverse Charge Mechanism is regulated by Article 199 of Council Directive 2006/112/EC on the common system of value-added tax (the VAT Directive). According to this directive, the Reverse Charge Mechanism can be applied for the supply of the following goods and services:

1. <u>Construction work or cleaning services.</u>

2. <u>Immovable property.</u>

3. <u>Transfer of allowances to emit greenhouse gases.</u>

4. Mobile phones.

5. Integrated circuit devices.

6. Gas and electricity.

7. <u>Gas and electricity certificates.</u>

8. <u>Telecommunication services</u>

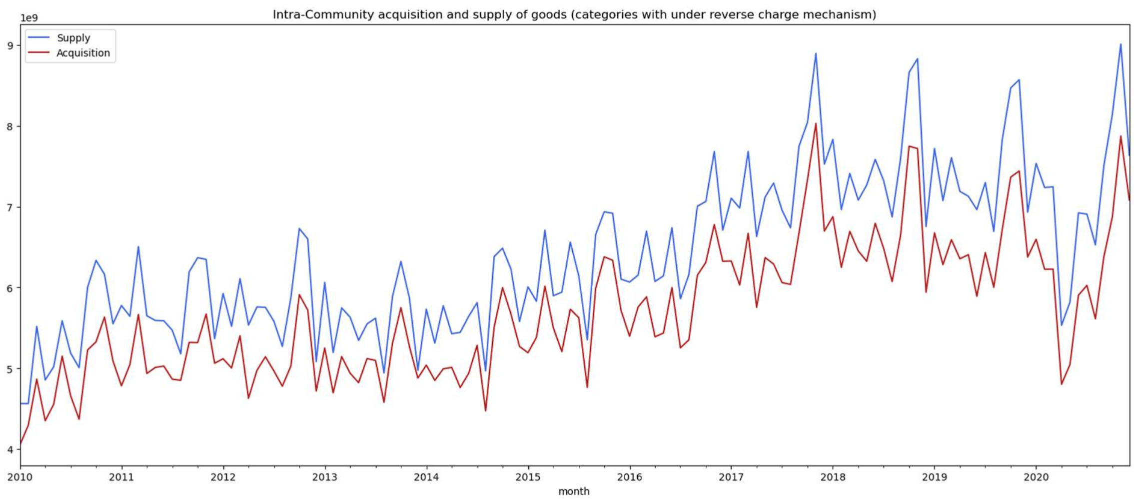
9. Game consoles, tablet PC's and laptops

10. Cereals and industrial crops not normally used for final consumption.

11. Raw and semi-finished metals.

As mentioned earlier, several of the products and services that are under Reverse Charge Mechanism are not included in the Intrastat dataset (those are the items underlined in the list above), as it is mainly focused on movable goods. It is also important to note that the categories under Reverse Charge Mechanism and those present in the analysed data do not always perfectly correspond – some categories subject to Reverse Charge Mechanism are defined on a lower level than the 4-digit CN classification.[104] In those cases the whole category was used, regardless of what share of it falls under the Reverse Charge Mechanism. The discrepancy between Intra-Community Supply and Intra-Community Acquisition is visibly higher (see Figure 11) than that observed for all categories covered by the Intrastat. The total surplus of Intra-Community Supply over Intra-Community Acquisition amounts to about 12.6% of intra-Community Supply.

Figure 11: Declared export (ICS) and import (ICA) for categories under Reverse Charge Mechanism between the seven selected Member States (Germany, France, Poland, the Netherlands, Hungary, Latvia and Malta), 2010-2020



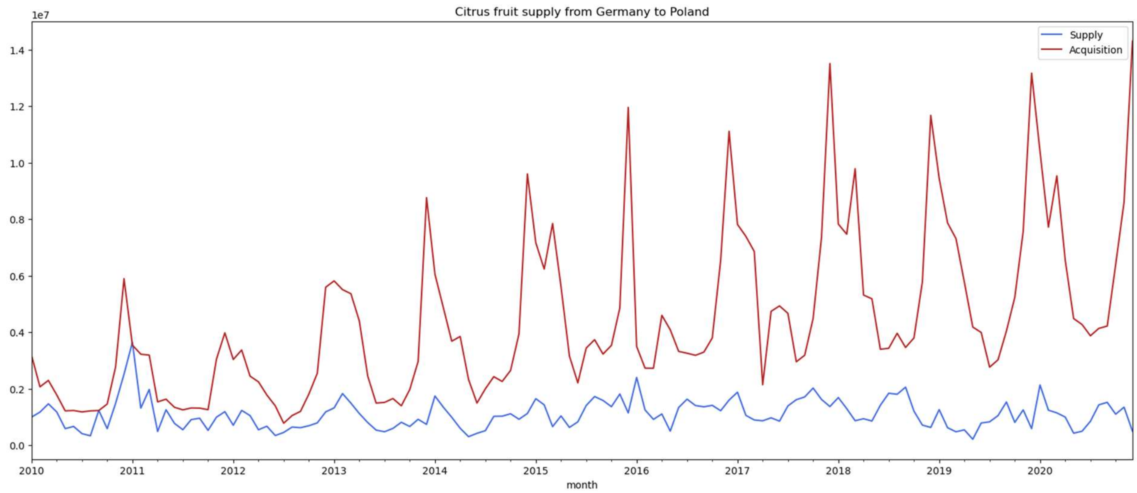
Source: own elaboration based on Intrastat EU trade since 1988 by HS2-4-6 and the CN8 dataset

It is important to stress that the size of the discrepancy cannot be simply interpreted as the size of tax fraud (be it MTIC or any other type). As mentioned above, there can be a number of valid reasons for mirror statistics not converging, especially on the level of single categories, where the data is very noisy. The general observations are not always confirmed on a lower scale. Figure 12 illustrates one such example – the case of Citrus fruit trade flows declared by Germany and Poland, where the value of Intra-Community Supply greatly exceeds the value of Intra-Community Acquisition. On top of that, the overall pattern seems somewhat different – the time series for acquisitions shows a very high level of seasonality, while the supply is much less volatile, which was likely caused by the fact that seasonal suppliers are below the Intrastat registration threshold. This example already proves that a more sophisticated method of anomaly detection is necessary to identify the type of discrepancy that can be

[104] The list of CN4 codes used: 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1201, 1202, 1203, 1204, 1205, 1206, 1207, 1208, 1209, 1210, 1211, 1212, 1213, 1214, 7101, 7102, 7103, 7104, 7105, 7106, 7107, 7108, 7109, 7110, 7111, 7112, 7114, 7115, 7116, 7117, 7204, 7301, 7302, 7303, 7304, 7305, 7306, 7307, 7308, 7309, 7310, 7311, 7312, 7313, 7314, 7315, 7316, 7317, 7318, 7319, 7320, 7321, 7322, 7323, 7324, 7325, 7326, 7404, 7411, 7412, 7413, 7415, 7418, 7419, 7503, 7507, 7508, 7602, 7608, 7609, 7610, 7611, 7612, 7613, 7614, 7615, 7616, 7802, 7806, 7902, 7907, 8002, 8007, 8101, 8102, 8103, 8104, 8105, 8106, 8107, 8108, 8109, 8110, 8111, 8112, 8113, 8201, 8202, 8203, 8204, 8205, 8206, 8207, 8208, 8209, 8210, 8211, 8212, 8213, 8214, 8215, 8301, 8302, 8303, 8304, 8305, 8306, 8307, 8308, 8309, 8310, 8311, 8443, 8517, 8525, 8527, 8528, 8542, 8548, 9028, 9504.

associated with MTIC fraud; a simple comparison of trade mirror statistics (although on a large-scale indicative of some systematic underreporting on the side of Acquirers) does not work on a lower level.

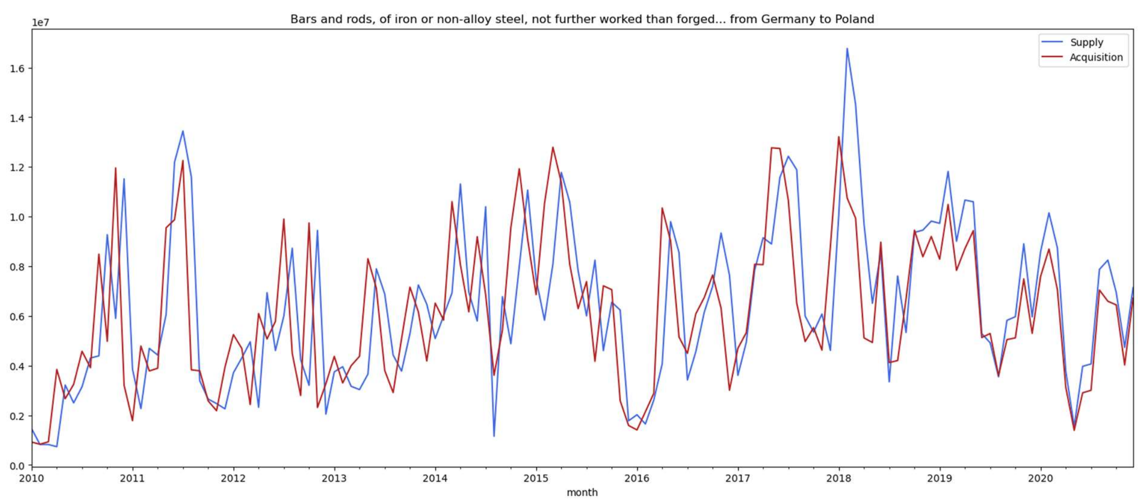
Figure 12: Declared export (ICS) and import (ICA) of "Citrus fruit" between Germany (supplier) and Poland (acquirer), 2010-2020



Source: own elaboration based on Intrastat EU trade since 1988 by HS2-4-6 and the CN8 dataset

Another example of an unexpected relationship between the values of declared Intra-Community Acquisition and Intra-Community Supply is found in the volumes of trade in *Bars and rods, of iron or non-alloy steel, not further worked than forged, hot-rolled, hot-drawn or hot-extruded, but incl. those twisted after rolling* between Germany and Poland (see Figure 13). In this case there is a clearly visible time shift pattern, where acquisition precedes supply (at least until the start of 2019) by one month. Even though the time lag is one of the known reasons for discrepancies, we would expect it to go the other way around, with the supply being reported before the acquisition.

Figure 13: Declared export (ICS) and import (ICA) of "Bars and rods, of iron or non-alloy steel, not further worked than forged, hot-rolled, hot-drawn or hot-extruded, but incl. those twisted after rolling" between Germany (supplier) and Poland (acquirer), 2010-2020



Source: Own elaboration based on Intrastat EU trade since 1988 by HS2-4-6 and the CN8 dataset

In conclusion, if a method based on international trade mirror statistics is used, many such issues will have to be considered when designing the final form of the algorithm. One of the key tasks when designing such an algorithm will be to identify the statistical patterns associated with tax fraud, as opposed to other possible factors contributing to the discrepancies. Because in reality some of the discrepancies are negative and some positive (as shown by the citrus example), they cancel each other out when calculating the total surplus of Intra-Community Supply over Acquisition. When calculating the total value of discrepancies (both ways) on a product level the value of discrepancies is tripled. The value would be even greater when calculated separately for country pairs. The lower we go down into the aggregation level the more noise is introduced, and the picture becomes more fuzzy.

VIES and VAT returns

One of the promising methods that was proposed by the Fiscalis Project Group is the use of matched information from VIES system and VAT returns in order to acquire discrepancies between those two registers. Because the access to the data from EC Sales Lists in VIES system is restricted to national administrations, most of the insights and conclusions on feasibility of that method is based on the knowledge of external experts, who had the chance to work with those registers.

VIES system contains trade data for single entity but aggregated to the period, so there is no information on particular transactions (and type of transacted good). Due to that there is a certain loss of information. On top of that, experts emphasized that the quality of the data is not always ideal and the use of this particular dataset might require significant effort on the part of data cleaning – this effort might even constitute a majority of necessary work for the described method. When it comes to the merger of the VIES data with VAT returns this should be fairly easy for Member State Administrations – both registers contain the *id* in the form of VAT number. If the data was shared with the contractor, VAT numbers would likely need to be substituted with some other unique id numbers.

The use of VIES system for the purpose of MTIC fraud detection (and estimation of the size) is not strictly limited to the method based on matching it with VAT returns – the absence of the taxpayer in that register altogether can also be used as a predictor for the fraud. According to one of the experts, among identified fraudsters (strictly speaking only missing traders themselves), only around 17% were registered in VIES. It's unclear whether this is the case after introduction of the quick fixes in 2020, but this is most likely true for years prior to their introduction, limiting time coverage of the method based on matching with VAT returns.

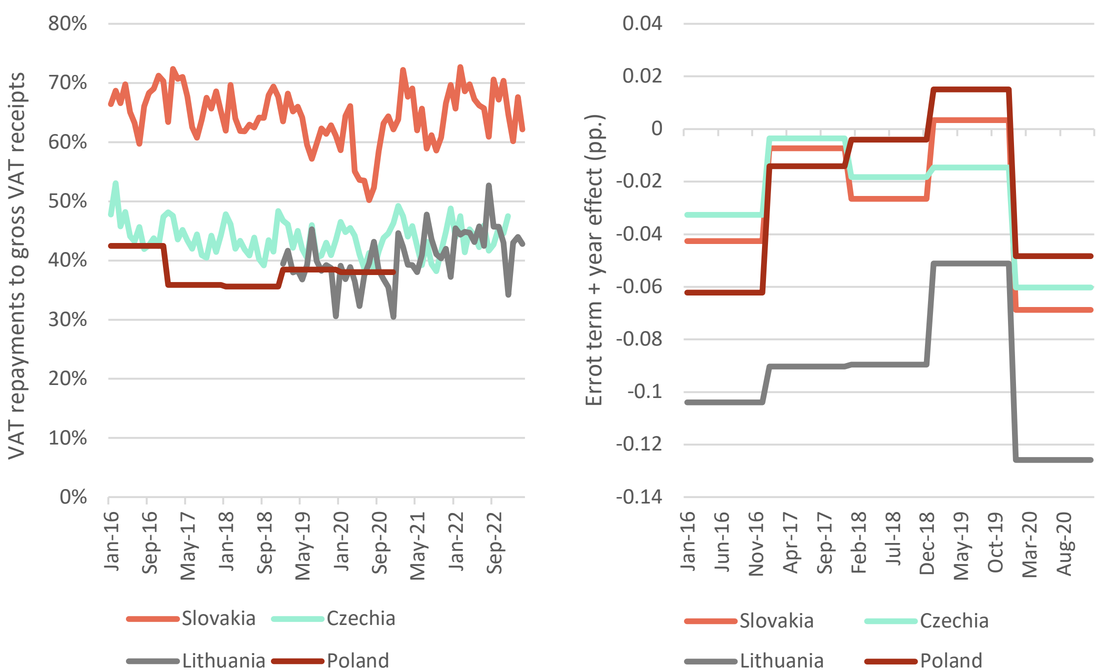
Importantly, as discussed in Section IV.b, the missing trader may report their intra-Community acquisition in order to obfuscate the identification of fraud (see Figure 8). In this case, the comparison of data in VIES and VAT returns would not be informative. At the same time, reporting the intra-Community acquisition leaves another trace and creates the possibility of matching them with the following domestic transactions. Cases where no such transactions are recorded signal fraudulent activity. It needs to be noted that the need to match VAT returns data could have an impact on the precision of the estimates. Tracking the fraud in intermediate goods requires controlling for the possible transformation of goods and services acquired by legitimate traders. Since there are many possibilities for transforming intermediate goods down the supply chain, this kind of matching would be prone to some error.

VAT repayments and unexplained dynamics of the EU-wide VAT compliance gap

To verify whether VAT repayments and unexplained dynamics of the EU-wide VAT compliance gap could be used as partial or secondary indicators of MTIC fraud, the study team has compiled these statistics for four Member States (Czechia, Lithuania, Poland and Slovakia) for which data on VAT repayments was available (see Figure 14).

As visualised by the graphs, there is no relationship visible at first glance. Correlation between all the indicators is very low and not significant statistically. Although this is not a conclusive observation, it may be showing that the partial indicators of the fraud include large noise, i.e., are largely affected by other factors. For this reason, such indicators could likely only be used as secondary evidence.

Figure 14: VAT repayments and unexplained dynamics of the VAT compliance gap in selected Member States (2016-2023)



Source: own elaboration based on EC/CASE (2022). Note: only yearly data for VAT repayments for Poland was available.

Summary

As discussed in this chapter there are multiple sources of inaccuracies pertaining to analytical methods and data included under each methodological scenario. Table 11 summarizes these sources.

Table 11: Source of inaccuracies in different scenarios

Methodological scenario	Sampling, specification and estimation errors	Other sourced of inaccuracies
Scenario #1: Econometric analysis of trade fluctuations & simple analysis of mirror statistics	Specification and estimation errors of the statistical/econometric models.	Noise in the trade data related to different registration thresholds, size of companies across borders and discrepancies in the period of the declaration.
Scenario #2: Using classification algorithms	Specification and estimation errors of the statistical/econometric models.	Noise in the trade data related to different registration thresholds, size of companies across borders and discrepancies in the period of the declaration. Likely errors involved in the creation of the training dataset (wrong attribution of "fraudulent" and not fraudulent observations.
Scenario #3: Using clustering algorithms	Specification and estimation errors of the statistical/econometric models.	Noise in the trade data related to different registration thresholds, size of companies across borders and discrepancies in the period of the declaration.
Scenario #4: Discrepancies in VIES and VAT returns	-	Potential error/omissions of the modeller. Imprecision in describing potential changes to classification of processed products.
Scenario #5: Risk-based audit methods	Combination of the specification and estimation errors involved in modelling the selection for audits and modelling non-compliance ratio	Inherent omission of variables in the econometric models.
Scenario #6: Structural equation modelling	Sampling error underlying compilation of some indicators (e.g., GDP figures).	Noise in both primary and secondary indicators included in the modelling.
Scenario #7: Kalman filter	Sampling error underlying compilation of some indicators (e.g., GDP figures).	Noise in both primary and secondary indicators included in the modelling.

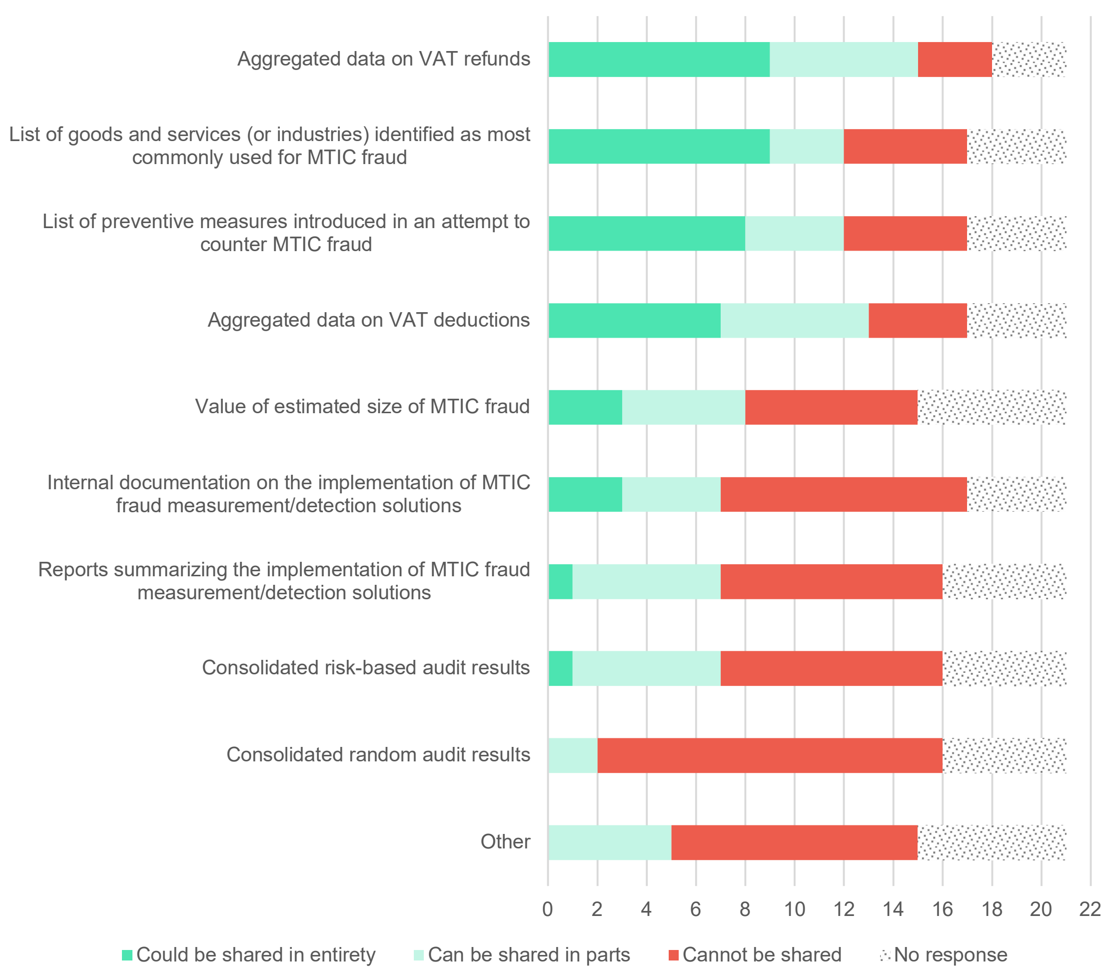
Source: own elaboration

V.b. Completeness

The completeness criterion refers to the extent to which each methodological scenario is able to cover all of the MTIC fraud types and types of transactions. Earlier, in Chapter I of this report, we distinguished two basic types of MTIC fraud – simple acquisition fraud and carousel fraud – as well as a number of tactics used to complicate transaction chains. For the purposes of evaluating completeness, this assessment considers two basic types of fraud and whether different tactics (such as contra-trading and cross-invoicing) could be captured by the methodological scenario in question.

Access to the necessary data is a crucial factor when assessing the feasibility of chosen methods. In the questionnaire, respondents were asked about their administrations' willingness to share relevant data with the study team. Figure 15 and Figure 16 summarize responses to two questions on access to data (the latter being a follow-up question specifically regarding VIES data). Here the results were even more pronounced than in the case of data availability alone. Of all the sources listed, none could be shared in full by more than half of the Member States and even upon expanding to include partial access, only a few passed this threshold. Those were: lists of goods, services and industries identified as at risk of MTIC fraud (12 of 21); lists of preventative measures introduced in an attempt to counter MTIC fraud (12 of 21); and aggregated data on VAT refunds (15 of 21) and on VAT deductions (13 of 21) – significantly overlapping with data sources that were previously also identified as the most available. Consolidated audit results (both random and risk-based) and reports summarizing the implementation of MTIC fraud measurement/detection solutions ranked the lowest, as for each of these sources no more than one Member State said they could be shared in their entirety. Furthermore, four of the Member States did not mark any of the requested data sources as available to be shared.

Figure 15: Summary of responses to questions on sources of information that can be shared by the Member States

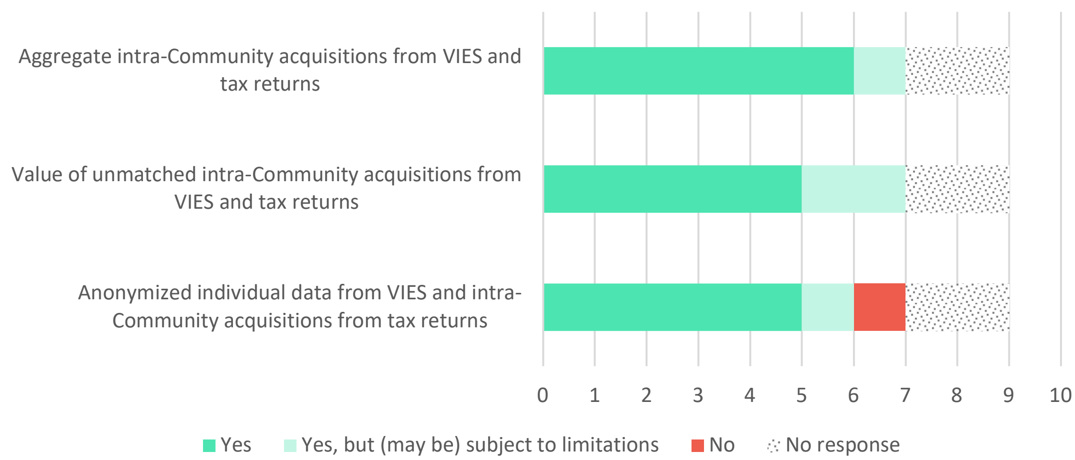


Source: own elaboration based on tax administration responses to questionnaire.

The second question, which was added to the amended version of the survey, asked respondents about the possibility of sharing data on Intra-Community Acquisitions from VIES and tax returns. A summary of answers to the question on the potential access to data on Intra-Community Acquisitions from VIES and tax returns for the study team, based on eight responses,[105] is presented Figure 16.

Overall, the majority of respondents declared that aggregate Intra-Community Acquisitions, the value of unmatched ones, and anonymised individual Intra-Community Acquisition data from VIES and tax returns could be made available, though in some cases with the caveat that limitations may apply.

Figure 16: Summary responses to follow-up question on additional sources that could be shared by Member States



Source: own elaboration based on tax administration responses to questionnaire.

Fluctuations in trade values and mirror statistics serve as the baseline indicators in scenarios #1- #3, and for validation purposes are further supplemented by irregularities in VAT repayments and unexplained variation in the overall VAT compliance gap. Since trade value data will only capture schemes where the commodities (or services) are moved repeatedly across borders, the analysis of their fluctuations is limited to carousel fraud. Mirror statistics, on the other hand, will likely capture all schemes, unless the supplier fails to register its transaction in Intrastat. VAT repayments will again only capture the schemes in which goods are moved continuously. However, brokers need to have little output VAT, which will likely not be the case in sophisticated schemes. Despite the limitations of these individual indicators, the first three scenarios are expected to capture all types of schemes, both simple and sophisticated, repetitive schemes.

Scenario #4 is expected to cover all types of MTIC fraud as in all or nearly all cases the basic assumptions of the method will be met (registration in VIES and non-registration in VAT returns, see Section V.a). Under scenario #5, the estimation would use the results of risk-based audits, which are likely to uncover the majority of fraud (though often post-factum), regardless of type. Yet, in case of more complicated fraud schemes, once can expect the accuracy of audit assessment to be lower and, rendering estimates using audit results somewhat inaccurate (depending on factors such as the exact type of audit or the level of expertise of the auditor). Scenarios #6 and #7 combine and synthetically

[105] However, conclusions are drawn based on seven responses, as one of the respondents left this section blank.

analyse multiple indicators of fraud from different data sources, likely allowing them to produce an estimate covering all types of MTIC fraud. Yet, some of the indicators may be ill-suited for covering simple acquisition fraud (fluctuation of trade values, VAT repayments) and complicated schemes (VAT repayments), which may decrease the accuracy of the entire methodological scenario in uncovering the scale of these types of fraud. A summary of the coverage of different schemes in each scenario is presented in Table 12 below.

Table 12: Completeness of proposed scenarios

Methodological scenario	Simple schemes (e.g., acquisition fraud)	Carousel fraud	Other complex schemes (e.g., contra-trading fraud, cross-invoicer fraud)	Other/comments
Scenario #1: Econometric analysis of trade fluctuations & simple analysis of mirror statistics	Yes (but not by all indicators included)	Yes	Yes (but not by all indicators included)	Weak coverage MTIC fraud in services
Scenario #2: Using classification algorithms	Yes (but not by all indicators included)	Yes	Yes (but not by all indicators included)	Weak coverage MTIC fraud in services
Scenario #3: Using clustering algorithms	Yes (but not by all indicators included)	Yes	Yes (but not by all indicators included)	Weak coverage MTIC fraud in services
Scenario #4: Discrepancies in VIES and VAT returns	Yes	Yes	Yes	-
Scenario #5: Risk-based audit methods	Yes	Yes	Yes (but likely lower accuracy)	-
Scenario #6: Structural equation modelling	Yes (but likely lower accuracy)	Yes	Yes (but likely lower accuracy)	Weak coverage MTIC fraud in services
Scenario #7: Kalman filter	Yes (but likely lower accuracy)	Yes	Yes (but likely lower accuracy)	Weak coverage MTIC fraud in services

Source: own elaboration

All of the proposed scenarios make use of methods requiring (to varying degrees) data from individual Member States' tax administrations. Member States' ability and willingness to share this data is therefore going to determine which of the methodological scenarios are actually viable and could be implemented. The key data sources which would need to be shared, in the context of the proposed scenarios, are: data on VAT refunds and deductions, consolidated risk-based audit results, and data on intra-Community acquisitions. As part of the questionnaire, Member States answered questions on the availability of data and their willingness to share the abovementioned sources. Although these answers do not necessarily guarantee that the study team will be able to obtain the data in question, they can be taken as an indicator. Table 13 combines this information with the data requirements for the different methodologies under consideration in order to show in how many Member States the respective scenarios could be employed.

Table 13: Expected coverage

Methodological scenario	Member State coverage based on data availability	Member State coverage based on data accessibility by contractor
Scenario #1: Econometric analysis of trade fluctuations & simple analysis of mirror statistics	27	27
Scenario #2: Using classification algorithms	27	27
Scenario #3: Using clustering algorithms	27	27
Scenario #4: Discrepancies in VIES and VAT returns	27	7-24 (expected ca. 16)
Scenario #5: Risk-based audit methods	11-21 (expected ca. 16)	7-15 (expected ca. 14)
Scenario #6: Structural equation modelling	27	27
Scenario #7: Kalman filter	27	27

Notee: Based on 21 responses (9 in the case of scenario #4). One of the Member States did not answer this question, and was therefore not counted as able to share, but did indicate all the data of interest was at the very least available.
Source: own elaboration.

As concerns the availability of the methodological scenarios based on administrative data, the information necessary to calculate discrepancies in VIES and VAT returns is in principle available to all Member States (despite some differences in the variables included). Although the construction of this option incorporates two alternative methods of execution to maximise potential data availability for the study team, it is still likely that that the information would not be made available. Similarly, Scenario #5 is unlikely to be feasible due to lack of access to risk-based audit results. Surprisingly, a number of respondents pointed to the lack of consolidated audit results, which means that the method is currently also not available for all Member State administrations.

V.c. Time covered

A summary of the extent to which each of the proposed scenarios is able to cover the required period is presented in Table 14. All but one out of the seven presented methodological scenarios build on datasets which are available for the period 2018-2022 (the ceiling set for the time coverage requirement) – namely the intra-Community trade figures reported in Intrastat (or EC sales returns) and in the national reporting obligations' system available in Eurostat's international trade in services series. Two scenarios are an exception to this. Scenario #4 relies on data from the EU VIES, which was introduced following the introduction of the single market in 1993. However, the data sufficient for the implementation of this scenario is only available from 2020. This year saw a legislative change which established the VAT identification number as one of the requirements to qualify for a zero VAT rate, providing strong incentive for its use. Scenario #5, on the other hand, uses risk-based audit methods and thus is reliant on access to national risk-based audit results. Based on the questionnaire answers presented in Chapter IV (Figure

15) we can see that, from the Member States which responded, around half is in possession of risk-based audit results (9 out of 21 have them for the entire period in question), which is a promising result. However, in practice this scenario is unlikely to be feasible, as the overwhelming majority of those countries declared that they were not in a position to share said results with the study team and so it is highly unlikely that a sufficient number of Member States would grant it.

Table 14: Time coverage

Methodological scenario	Period covered
Scenario #1: Econometric analysis of trade fluctuations & simple analysis of mirror statistics	2018-2022
Scenario #2: Using classification algorithms	2018-2022
Scenario #3: Using clustering algorithms	2018-2022
Scenario #4: Discrepancies in VIES and VAT returns	From 2020 onwards
Scenario #5: Risk-based audit methods	2018-2022
Scenario #6: Structural equation modelling	2018-2022
Scenario #7: Kalman filter	2018-2022

Source: own elaboration

V.d. Granularity

Table 15 presents the granularity of each scenario, which here refers to its ability to link the amount of VAT fraud tax gap to specific drivers and/or types of fraud. As was mentioned earlier in this chapter, intra-Community trade values, trade mirror statistics and VAT repayments are each likely to capture different types of fraud schemes, which in turn means that under scenarios #1, #2, and #3, estimates of revenue lost could be broken down by general type of scheme. However, it must be kept in mind that secondary indicators used under these approaches (unexplained variation of the compliance gap and VAT repayments) contain a lot of noise (impact of other factors than this of interest), which may hinder such a breakdown.

Under scenario #4, scrutiny of discrepancies between VIES and VAT returns, there is no possibility of detecting types of fraud. Although figures in VIES do not contain information on traded goods, the NACE codes of taxpayers from the merged datasets could be used indirectly to assign a broad category of traded goods and services.

Scenario #5 draws on risk-based audit results, which allow for a more detailed look into the characteristics of identified fraud cases. Audit data results are expected to contain both detailed information on goods trade, information on the role of taxpayer in a fraudulent scheme and type of scheme. Under scenarios #6 and #7 all of the endogenous variables will be in aggregate form, yielding a single overall estimate for a given Member State and time period.

Table 15: Granularity of the proposed scenarios

Methodological scenario	By a general type of scheme	By type of product
Scenario #1: Econometric analysis of trade fluctuations & simple analysis of mirror statistics	Yes, but limited accuracy	Yes, 4- or 6-digit CN codes (goods) and all service categories
Scenario #2: Using classification algorithms	Yes, but limited accuracy	Yes, 4- or 6-digit CN codes (goods) and all service categories
Scenario #3: Using clustering algorithms	Yes, but limited accuracy	Yes, 4- or 6-digit CN codes (goods) and all service categories
Scenario #4: Discrepancies in VIES and VAT returns	No	Yes, but indirectly and limited accuracy
Scenario #5: Risk-based audit methods	Yes	Yes
Scenario #6: Structural equation modelling	No	No
Scenario #7: Kalman filter	No	No

Source: own elaboration

V.e. Complexity and costs

The overall effort required to implement each of the methodological scenarios was assessed based on the SCM model with inputs to the model provided from various sources. Such a monetisation of effort is treated as a proxy of the complexity of different approaches as the calculation encompasses time (in the form of FTE) and competences of team members (in the form of fees) required.

For most of the scenarios, the calculation was carried out for 27 Member States rather than on a per-Member State basis, as the work on the approaches is not proportional to number of Member States covered. The processes are largely repeated, so there are large economies of scale. However, handling the data for 27 countries (meaning 351 trading pairs) requires different tools than would be required for a smaller coverage. For Scenarios #4 and #5, the effort was initially estimated per Member State, as the work here is proportional to number of Member States covered. 50% economies of scale were assumed. The calculation covers both single implementation of the methodological scenario (Table 16) and a single update (Table 17) that is expected to require from ca. 40% of the cost of the initial implementation. Regardless of the method implemented, the responses of interviewees pointed to ca. 40% of the cost of updating methodologies based on large datasets and individual-level figures (under the assumption of the same team implementing an update benefitting from the tools prepared with during the first implementation).

The calculation of effort required to apply the method is based on the experience of Member State administrations and the study team's own experience in implementing different analytical approaches to calculating tax compliance gaps. Evidence of the cost and complexity was obtained for the methods based on risk-based audit results, top-down estimates using national accounts and for estimating the PIT liability using survey data. The responses were averaged. Nevertheless, the study team decided to revise downwards the estimate of FTE required (from 20 to 40% depending on the approach). This downward revision was based on the Team's own experience in applying the same or similar methodologies by certain members of the study team.

The rates used for the calculation of the costs are based on Eurostat's mean annual earnings by sex, age and occupation available for all EU Member States for 2018.[106] As advised by Better Regulation Guidelines, a 25% overhead was assumed.[107] In every scenario, we assumed that the team comprises 10% managers, 40% experienced professionals, 25% associates and 25% junior staff.

Table 16: Estimated effort and cost for the implementation covering 5-year period

Scenario	Core tasks	First implementation for a 5-year period		
		FTE (monthly)	Total cost (27 MS)	Explanation and source of estimates
1	Econometric analysis of trade fluctuations	8	70 569	Source: own experience confirmed by the range of estimates of two administrations implementing hybrid approaches. Compilation of the database at 4-digit CN codes with additional figures expected to take 2 FTEs (included in the FTEs envisaged for the first two tasks)
	Simple analysis of mirror statistics	5		
	Simple analysis of irregularities in VAT repayment dynamics	2		
	Unexplained variation in the overall VAT compliance gap	1		
2	Classification algorithms to analyse trade statistics	17	92 622	Source: own experience confirmed by the range of estimates of two administrations implementing hybrid approaches. Compilation of the database at 4-digit CN codes with additional figures expected to take 2 FTEs (included in the FTEs envisaged for the first task). Additionally, 5x3 FTEs required for modelling with the use of machine learning techniques.
	Econometric analysis of irregularities in VAT repayment dynamics	3		
	Unexplained variation in the overall VAT compliance gap	1		
3	Clustering algorithms to analyse trade statistics	17	92 622	Source: own experience confirmed by the range of estimates of two administrations implementing hybrid approaches. Compilation of the database at 4-digit CN codes with additional figure expected to take 2 FTEs (included in the FTEs envisaged for the first task). Additionally, 5x3 FTEs required for modelling with the use of machine learning techniques.
	Econometric analysis of irregularities in VAT repayment dynamics	3		
	Unexplained variation in the overall VAT compliance gap	1		
4	Scrutiny of discrepancies in VIES and VAT returns	40.5	178 629	Assuming that the data received contains, individual-level taxpayer data from both

[106] NACE Rev. 2, B-S excluding O [EARN_SES18_28__custom_736081]

[107] See: https://commission.europa.eu/law/law-making-process/planning-and-proposing-law/better-regulation/better-regulation-guidelines-and-toolbox_en.

Scenario	Core tasks	First implementation for a 5-year period		
		FTE (monthly)	Total cost (27 MS)	Explanation and source of estimates
				VIES and VAT returns. It was assumed that cleaning and merging of the data for a single Member State would require 3 FTE. It was assumed that the cost of covering 27 Member State would be 50% lower than the sum of the costs borne for a single Member State (i.e., 50%*3*27)
5	Risk-based audit methods	121.5	535 887	On average, such methods required 3 months and 3 people working on time per country. We assume economies of scale and saving 50% of the time (that would be required for 27 teams covering from scratch 27 MS)
6	Structural equation modelling	20	88 212	Requires completing most of the processes under Scenario 1 (data compilation, calculation of the discrepancies in data), but additional effort for the econometric work must be envisaged (2 additional FTEs assumed)
7	Kalman filter	20	88 212	Requires completing most of the processes under Scenario 1 (data compilation, calculation of the discrepancies in data), but additional effort for the implementation of Kalman filter must be envisaged (4 additional FTEs assumed)

Source: own elaboration.

Table 17: Estimated effort and cost for an update covering a 1-year period

Scenario	Core tasks	Single update for a 5-year period		
		FTE (monthly)	Total cost (27 MS)	Explanation and source of estimates
1	Econometric analysis of trade fluctuations	6	28 228	Source: responses to the questionnaire. 50% of effort involved in the seminal estimation envisaged
2	Classification algorithms to analyse trade statistics	8	37 049	
3	Clustering algorithms to analyse trade statistics	8	37 049	
4	Scrutiny of discrepancies in VIES and VAT returns	16	71 452	
5	Risk-based audit methods	49	214 355	
6	Structural equation modelling	8	35 285	
7	Kalman filter	8	35 285	

Source: own elaboration.

VI. Comparison

VI.a. Main results

This chapter presents a comparison of the methodological scenarios. The scenarios were compared using central values of the intervals for weights presented in Table 6. The sensitivity of the selection with respect to the weights assigned to each of the criteria is presented in the following Section VI.b.

The assessment (see Table 18 and Table 19) takes into account all the criteria discussed in Table 6 except for the estimated cost of the implementation. The objective of this assessment was to rank suitability of approaches regardless the effort involved.

This comparison indicates Scenario #2: *Using classification algorithms* and Scenario #3: *Using clustering algorithm* would likely be optimal (score of 0.82). Both scenarios are based on similar analytical methods and indicators used. Thus, at this stage it was not possible to differentiate the assessment for these two similar approaches. For this reason, both approaches were tested with some more insight on suitability of both discussed in the following Chapter VII.

Scenario #4: *Discrepancies in VIES and VAT returns* scored nearly as well as Scenario #2 and #3 (0.78 if implemented by the study team and 0.81 if implemented by administrations). Keeping in mind inevitable margin of error in the assessment, Scenario #4 should be regarded as equally promising. Scenario #4 appears to be the most complete and most accurate of all the scenarios. Yet, in contrast to all other scenarios, it could be used to cover only a short period of time (from 2020). It also does not allow for detailed breakdowns of the MTIC (as Scenario #5: *Risk-based audit methods* does).

The remaining scenarios (scenario #1 and scenarios #6-#8) obtained nearly identical scores (0.68-0.7) despite their different characteristics and scores assigned for various criteria. Somewhat lower score of these is driven by inability to break down fraud by type (for scenario #7 and #8), lower expected coverage (for scenario #6) and expected lower accuracy (for scenario #1).

Table 18: Assessment table (1)

Criteria	Subcriteria	Weight and method of scaling	**Scenario #1:** Econometric analysis of trade fluctuations & simple analysis of mirror statistics — Assessment (description)	Value	**Scenario #2:** Using classification algorithms — Assessment (description)	Value	**Scenario #3:** Using clustering algorithms — Assessment (description)	Value	**Scenario #4:** Discrepancies in VIES and VAT returns — Assessment (description)	Value
Accuracy, completeness, and comparability	Accuracy of point estimates/comparability across Member States	Weight in the overall comparison: 20% Floor: 5 pp. deviation on average Ceiling: expected full accuracy	The analysis of trade statistics point to significant noise in the data as there are different registration thresholds, size of companies across borders and discrepancies in the period of the declaration. Moreover, not all fraudulent transactions must necessarily be reported by the supplier. Using the discrepancy between ICS and ICA for goods unlikely to be subject to fraud, ca. 5 pp. deviation is expected	0	Using such algorithms allows to control for the noise in the data that is not controlled for under Scenario #1. We assume that 60% of the noise would be filtered reducing the error from 5 pp. to 1.5 pp.	0.7	Using such algorithms allows to control for the noise in the data that is not controlled for under Scenario #1. We assume that 60% of the noise would be filtered reducing the error from 5 pp. to 1.5 pp.	0.7	Interviews with expert point to marginal issues and nearly ideal accuracy of this approach	1
	Completeness across types of MTIC fraud (directly interrelated with the above)	Weight in the overall comparison: 12.5% Floor: two-thirds covered Ceiling: expected full coverage	Weaker coverage of fraud in services	0.8	Weaker coverage of fraud in services	0.8	Weaker coverage of fraud in services	0.8	Full completeness expected	1

			Scenario #1: Econometric analysis of trade fluctuations & simple analysis of mirror statistics		**Scenario #2:** Using classification algorithms		**Scenario #3:** Using clustering algorithms		**Scenario #4:** Discrepancies in VIES and VAT returns	
Criteria	**Subcriteria**	**Weight and method of scaling**	**Assessment (description)**	**Value**	**Assessment (description)**	**Value**	**Assessment (description)**	**Value**	**Assessment (description)**	**Value**
	Comparability across time/accuracy of trends	Weight in the overall comparison: 17.5% Floor: 10% of observations classified as structural breaks Ceiling: expected full accuracy	No structural breaks expected	1	No structural breaks expected	1	No structural breaks expected	1	No structural breaks expected	1
	Coverage of Member States/level of extrapolation needed to achieve coverage across EU-27	Weight in the overall comparison: 15% Floor – no coverage Ceiling –: full coverage	All 27 Member State expected to be covered	1	All 27 Member State expected to be covered	1	All 27 Member State expected to be covered	1	Expected 16 Member States covered if the calculations are done by the study team, all 27 Member States if implemented by administrations	**1**/0.59
	Coverage of time/timeliness	Weight in the overall comparison: 15% Floor – single year from 2018-2022 covered. Ceiling – 2018-2022	2018-2022	1	2018-2022	1	2018-2022	1	From 2020-onwards	0.6
Granularity	Ability to link the amount of the VAT fraud tax gap to specific drivers/types of fraud	Weight in the overall comparison: 20% 0: No breakdown 0.5: Possibility of breakdown by types of irregularities or types of taxpayers or types of goods 1: Possibility of breakdown by types of irregularities, types of goods and types of taxpayers	Possibility of breakdown by basic types of irregularities and types of goods	0.5	Possibility of breakdown by basic types of irregularities and types of goods	0.5	Possibility of breakdown by basic types of irregularities and types of goods	0.5	Indirect breakdown by type of products	0.5

Criteria	Subcriteria	Weight and method of scaling	**Scenario #1:** Econometric analysis of trade fluctuations & simple analysis of mirror statistics — Assessment (description)	Value	**Scenario #2:** Using classification algorithms — Assessment (description)	Value	**Scenario #3:** Using clustering algorithms — Assessment (description)	Value	**Scenario #4:** Discrepancies in VIES and VAT returns — Assessment (description)	Value
		Assessment	**0.68**		**0.82**		**0.82**		**0.78 – implementation by the study team** **0.81 – implementation by Member State administrations**	

Source: own elaboration.

Table 19: Assessment table (2)

Criteria	Subcriteria	Weight and method of scaling	Scenario #5: Risk-based audit methods – Assessment (description)	Scenario #5: Value	Scenario #6: Structural equation modelling – Assessment (description)	Scenario #6: Value	Scenario #7: Kalman filter – Assessment (description)	Scenario #7: Value
Accuracy, completeness, and comparability	Accuracy of point estimates/comparability across Member States	Weight in the overall comparison: 20% Floor: 5 pp. deviation on average Ceiling: expected full accuracy	This methodological scenario depends to large extent on the effectiveness of the audit function. Moreover, the source of inaccuracies in the case of risk-based audits is the combination of the statistical errors involved in modelling the selection for audits and modelling non-compliance ratio. The size of these errors depends on the number of audits performed, non-compliance rate and accuracy of audit assessment. Unfortunately, the components of this error are not observed by the study team. Yet, expected accuracy of the approach of the Member State administrations is relatively high – higher than in the case of methodologies under Scenario #2 and #3. Thus, we assume that the error could be lower than the error expected for Scenario #2 and #3.	0.8	As shown in the accuracy assessment, there is likely a large noise in the secondary indicators included under Scenario #6 and #7. It could be expected that that the inaccuracies of this scenario could be in the range of the inaccuracies of the MIMIC approach for estimating the underground economy. Assuming that the inaccuracy of the approach is around the mean deviation of estimates in the studies using the MIMIC approach and lower ratio of the MTIC fraud (compared to the relative share of the underground economy), we expect ca. 2.5 pp deviation.	0.5	As shown in the accuracy assessment, there is likely a large noise in the secondary indicators included under Scenario #6 and #7. It could be expected that that the inaccuracies of this scenario could be in the range of the inaccuracies of the MIMIC approach for estimating the underground economy. Assuming that the inaccuracy of the approach is around the mean deviation of estimates in the studies using the MIMIC approach and lower ratio of the MTIC fraud (compared to the relative share of the underground economy), we expect ca. 2.5 pp deviation.	0.5
	Completeness across types of MTIC fraud (directly interrelated with the above)	Weight in the overall comparison: 12.5% Floor: two-thirds covered Ceiling: expected full coverage	Full completeness expected	1	Weaker coverage of fraud in services	0.8	Weaker coverage of fraud in services	0.8

			Scenario #5: Risk-based audit methods		**Scenario #6:** Structural equation modelling		**Scenario #7:** Kalman filter	
Criteria	**Subcriteria**	**Weight and method of scaling**	**Assessment (description)**	**Value**	**Assessment (description)**	**Value**	**Assessment (description)**	**Value**
	Comparability across time/accuracy of trends	Weight in the overall comparison: 17.5% Floor: 10% of observations classified as structural breaks Ceiling: expected full accuracy	Reforms of audit procedures could significantly impact the accuracy of trend estimates	0	No structural breaks expected	1	No structural breaks expected	1
	Coverage of Member States/level of extrapolation needed to achieve coverage across EU-27	Weight in the overall comparison: 15% 0: no coverage 1: full coverage	Expected ca. 11 Member State to be covered if implemented by the study team, expected ca. 11 Member State to be covered if implemented by Member State administrations	**0.41**/0.59	All 27 Member State expected to be covered	1	All 27 Member State expected to be covered	1
	Coverage of time/timeliness	Weight in the overall comparison: 15% Floor –single year from 2018-2022 covered. Ceiling – 2018-2022	2018-2022	1	2018-2022	1	2018-2022	1

			Scenario #5: Risk-based audit methods		Scenario #6: Structural equation modelling		Scenario #7: Kalman filter	
Criteria	**Subcriteria**	**Weight and method of scaling**	**Assessment (description)**	**Value**	**Assessment (description)**	**Value**	**Assessment (description)**	**Value**
Granularity	Ability to link the amount of the VAT fraud tax gap to specific drivers/types of fraud	Weight in the overall comparison: 20% 0: No breakdown 0.5: Possibility of breakdown by types of irregularities or types of taxpayers or types of goods 1: Possibility of breakdown by types of irregularities, types of goods and types of taxpayers	Possibility of breakdown by types of irregularities and types of taxpayers	1	No breakdown	0	No breakdown	0
		Assessment	**0.7 – implementation by the study team 0.73 – implementation by administrations**		**0.68**		**0.68**	

Source: own elaboration.

VI.b. Sensitivity check

As indicated in Section VI.e., the Monte Carlo simulation consists of drawing the weights 10 000 times from uniform distributions in the restricted intervals for the six assessment criteria defined in Table 6. For each of the iterations, we have computed the simulated performance of the seven methodological scenarios (A-G) proposed for calculating the MTIC gap. To rank the methods, we used dominance relations proposed by Mazurek and Strzałka (2022). These include: (1) comparing the means of the simulated indicators, (2) comparing the number of times when a certain method is the best one among those tested, and (3) pairwise comparisons of the number of times one method is superior to another method.

The basic indicators show the superiority of the methodological scenario #2 and #3 under assumed mean distributions of weights (Table 20). The average composite performance indicator is statistically significantly higher than any other considered method. Moreover, this indicator was superior to all other indicators in nearly 98% of comparisons made. below shows the simulated distributions of performance of each method. The density plots show a small degree of overlap between the empirical distribution of methodological scenario #2 and #3, and the next best scenario, #4 (see Figure 17). The empirical cumulative distribution presented in Figure 18 shows the dominance of methodological #2 and #3 to all other distributions.

The ranking of all the methods based on pairwise dominance can be derived from analysing the number of occurrences where a particular method was superior to another method (see Table 21). Methodological scenario #2 and #3 dominated all other methods in the vast majority of simulated iterations. Yet, in around 2.5% of cases of the assumed weights, the scenario #4 would be selected over #2 or #3.

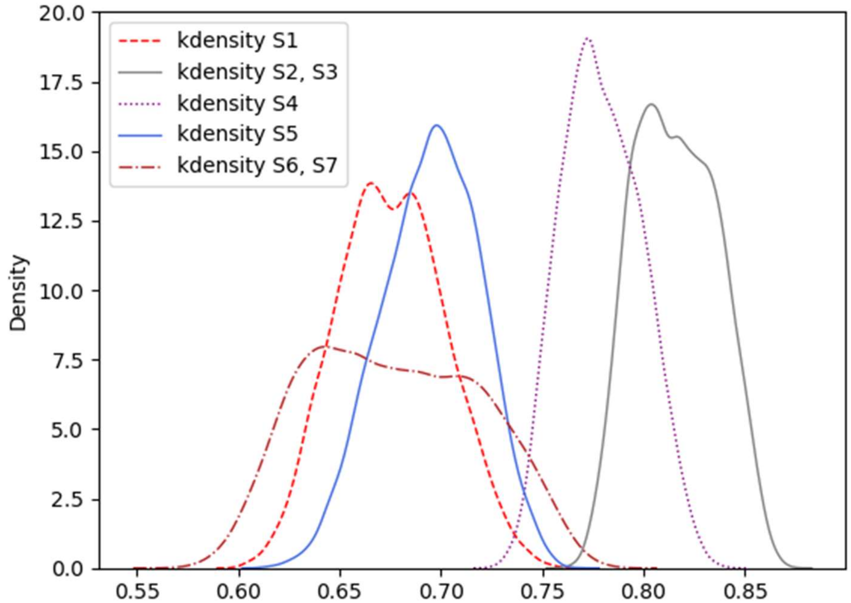
This analysed output from the Monte Carlo simulation depict strong dominance of three methodological scenarios, #2, #3 and #4, over other options. Nearly all combinations of weights that are close to the assumed means support the choice of scenario #2 or #3. Yet, there is some uncertainty regarding the selection of the methodological scenario #2 or #3, which scored best, over scenario #4. At the same time the assessment and comparison of scores, shows more clearly that scenarios #1, #5, #6 and #7 are expected to be less suitable for the objectives of the second phase of this study.

Table 20: Simulated performance indicators

Measure/Scenario	1	2	3	4	5	6	7
Mean	0.68	0.82	0.82	0.78	0.70	0.68	0.68
Standard error	0.03	0.02	0.02	0.02	0.02	0.04	0.04
Number of times best	0	9 756	9 756	244	0	0	0

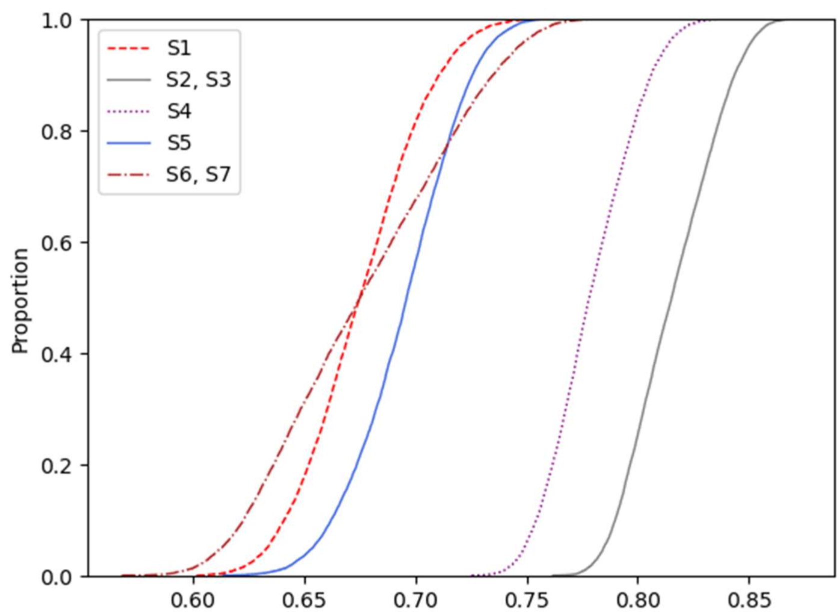
Source: own elaboration based on simulated values.

Figure 17: Simulated density of performance per method



Source: own elaboration based on simulated values.

Figure 18: Cumulative probability per method



Source: own elaboration based on simulated values.

Table 21: Pairwise dominance

Method	1	2	3	4	5	6	7
1	.	0	0	0	3 233	4 988	4 988
2	10 000	.	0	9 756	10 000	10 000	10 000
3	10 000	0	.	9 756	10 000	10 000	10 000
4	10 000	244	244	.	9 962	9 997	9 997
5	6 767	0	0	38	.	6 037	6 037
6	5 012	0	0	3	3 963	.	0
7	5 012	0	0	3	3 963	0	

Source: own elaboration based on simulated values.
Note: the table shows how many times the method listed in each row of the table was superior based on weighted average of performance indicators to the method listed in the column.

VII. Experimental implementation of Scenario #2 and #3

VII.a. Scope of the analysis

This chapter discusses the results of the experimental partial implementation of the two most promising methodological scenarios (*Scenario #2: Using classification algorithms* and *Scenario #3: Using clustering algorithms*), as indicated by the assessment. The objective of this implementation was to further reduce the risks related to the selection of the best-suited approach, improve the chances of success of these scenarios and pinpoint additional problems that could not be foreseen before such an experiment. In connection with this, the following sections focus on the patterns observed in the data, discuss measures of fit of different models and present the results of the simulation to verify whether the scale, trend and distribution of the MTIC gap is consistent with the expectations and with other sources.

It shall be noted that due to time constraints on the Phase I, the analysis remains partial. The analysis covers seven Member States[108] and only trade in goods (excluding services). Since the analysis covers trade between 42 country-pairs (7x6), it does not allow to calculate the entire value of the MTIC fraud in each country. The estimated value of irregularities could only be associated with the Intra-Community Supply and Intra-Community Acquisition within the analysed group of Member States.

It should also be noted that the implementation of the method for all Member States and groups of products and services will be much more resource intensive, involving thorough analysis to a number of methodological choices that need to be made, such as the choice of the model, variables, filtering,

[108] France, Germany, Italy, the Netherlands, Poland, Spain and the United Kingdom.

parameters of the model, granularity of the dataset and many more. The full analysis will also require the implementation of a modelling algorithm using different tools, able to handle this much larger dataset.

VII.b. Training dataset

To carry out the experimental implementation, described in the sections which follow, the study team constructed a training dataset. This training set consists of all the information needed for the purpose of supervised training or, in other words, estimating the models. Thus, in contrast to the entire dataset included in the analysis, all the observations in the dataset were classified as either *fraudulent* or *non-fraudulent*, a process described in more detail further in this section.

The initial dataset contains Intrastat information on Intra-Community Acquisition (*import*) and Intra-Community Sales (*export*), covering 41 product categories (in accordance with the Combined Nomenclature and at the 4-digit level), listed in Table 22, and spanning three countries: Germany, the Netherlands, and Poland.[109] The choice of countries (which were a subset of the countries covered in the larger dataset used in the experiments) was motivated by two aspects: the relative completeness of their trade data and the ease with which the study team could access reliable and detailed information on any changes in domestic reverse charge legislation, which was crucial for the next step of the process: creating the binary variable *rcm_import.* This variable was based on a review of publicly available legislative documents of each of the three countries and described which product categories were covered by the domestic reverse charge mechanism and when (see Table 22). It took the value 1 if a given product was covered by the domestic reverse charge mechanism in the importing country on a given month and 0 otherwise.

In order to ensure accurate training of the models, the dataset was designed to be as representative as possible in terms of the number of observations and variety of goods covered. This variety extended not only to the product categories themselves, but also their degree of susceptibility to MTIC fraud – the dataset included both categories which are almost certain to not be the subject of fraud (such as highly perishable foods, e.g., strawberries), categories which are known for often being targeted by fraudsters (e.g., mobile phones) and those which are not as clear cut and, i.e., have not been implicated in known fraud cases, but could nevertheless have been targeted, for instance sound recording equipment. This was done to guarantee that the dataset was not only representative but also had clear cases that the algorithm could "learn" from.

Table 22: Products included in the training dataset

Product code	Description	Domestic reverse charge mechanism introduction
102	Live bovine animals	
602	Live plants incl. their roots, cuttings and slips; mushroom spawn (excl. bulbs, tubers, tuberous roots, corms, crowns and rhizomes, and chicory plants and roots)	
702	Tomatoes, fresh or chilled	

[109] Limiting the training dataset to 41 products and three countries was optimal considering the tight time frame for this step, as even this smaller dataset would be sufficient for the purpose of training.

Product code	Description	Domestic reverse charge mechanism introduction
805	Citrus fruit, fresh or dried	
2702	Lignite, whether or not agglomerated (excl. jet)	Poland: 01/10/2013
7205	Granules and powders of pig iron, spiegeleisen, iron or steel (excl. granules and powders of ferro-alloys, turnings and filings of iron or steel, radioactive iron powders "isotopes" and certain low-calibre, substandard balls for ballbearings)	Germany: 01/10/2014 Poland: 01/10/2013
7213	Bars and rods of iron or non-alloy steel, hot-rolled, in irregularly wound coils	Germany: 01/10/2014 Poland: 01/10/2013
7214	Bars and rods, of iron or non-alloy steel, not further worked than forged, hot-rolled, hot-drawn or hot-extruded, but incl. those twisted after rolling (excl. in irregularly wound coils)	Germany: 01/10/2014 Poland: 01/10/2013
7215	Bars and rods, of iron or non-alloy steel, cold-formed or cold-finished, whether or not further worked, or hot-formed and further worked, n.e.s.	Germany: 01/10/2014 Poland: 01/10/2013
8517	Telephone sets, incl. telephones for cellular networks or for other wireless networks; other apparatus for the transmission or reception of voice, images or other data, incl. apparatus for communication in a wired or wireless network [such as a local or wide area network]; parts thereof (excl. than transmission or reception apparatus of heading 8443, 8525, 8527 or 8528)	Germany: 01/07/2011 Netherlands: 01/04/2013 Poland: 01/07/2015
406	Cheese and curd	
407	Birds' eggs, in shell, fresh, preserved or cooked	
408	Birds' eggs, not in shell, and egg yolks, fresh, dried, cooked by steaming or by boiling in water, moulded, frozen or otherwise preserved, whether or not containing added sugar or other sweetening matter	
409	Natural honey	
806	Grapes, fresh or dried	
807	Melons, incl. watermelons, and papaws "papayas", fresh	
808	Apples, pears and quinces, fresh	
809	Apricots, cherries, peaches incl. nectarines, plums and sloes, fresh	
810	Fresh strawberries, raspberries, blackberries, back, white or red currants, gooseberries and other edible fruits (excl. nuts, bananas, dates, figs, pineapples, avocados, guavas, mangoes, mangosteens, papaws "papayas", citrus fruit, grapes, melons, apples, pears, quinces, apricots, cherries, peaches, plums and sloes)	
1002	Rye	

Product code	Description	Domestic reverse charge mechanism introduction
1003	Barley	
1004	Oats	
1005	Maize or corn	
1006	Rice	
7201	Pig iron and spiegeleisen, in pigs, blocks or other primary forms	Germany: 01/10/2014 - present
7202	Ferro-alloys	Germany: 01/10/2014 - present Poland: 01/10/2013 - present
7203	Ferrous products obtained by direct reduction of iron ore and other spongy ferrous products, in lumps, pellets or similar forms; iron having a minimum purity by weight of 99,94%, in lumps, pellets or similar forms	
7204	Ferrous waste and scrap; remelting scrap ingots of iron or steel (excl. slag, scale and other waste from the production of iron or steel; radioactive waste and scrap; fragments of pigs, blocks or other primary forms of pig iron or spiegeleisen)	Germany: 01/10/2014 - present Netherlands: 01/01/2007 - present Poland: 01/10/2013 - present
7206	Iron and non-alloy steel in ingots or other primary forms (excl. remelting scrap ingots, products obtained by continuous casting and iron of heading 7203)	Germany: 01/10/2014 - present
7208	Flat-rolled products of iron or non-alloy steel, of a width >= 600 mm, hot-rolled, not clad, plated or coated	Germany: 01/10/2014 - present Poland: 01/10/2013 - present
7209	Flat-rolled products of iron or non-alloy steel, of a width of >= 600 mm, cold-rolled "cold-reduced", not clad, plated or coated	Germany: 01/10/2014 - present Poland: 01/10/2013 - present
7210	Flat-rolled products of iron or non-alloy steel, of a width >= 600 mm, hot-rolled or cold-rolled "cold-reduced", clad, plated or coated	Germany: 01/10/2014 - present Poland: 01/10/2013 - present
7211	Flat-rolled products of iron or non-alloy steel, of a width of < 600 mm, hot-rolled or cold-rolled "cold-reduced", not clad, plated or coated	Germany: 01/10/2014 - present Poland: 01/10/2013 - present
7212	Flat-rolled products of iron or non-alloy steel, of a width of < 600 mm, hot-rolled or cold-rolled "cold-reduced", clad, plated or coated	Germany: 01/10/2014 - present

Product code	Description	Domestic reverse charge mechanism introduction
		Poland: 01/10/2013 - present
7222	Other bars and rods of stainless steel; angles, shapes and sections of stainless steel, n.e.s.	Germany: 01/10/2014 - present
8518	Microphones and stands therefor (excl. cordless microphones with built-in transmitter); loudspeakers, whether or not mounted in their enclosures; headphones and earphones, whether or not combined with a microphone, and sets consisting of a microphone and one or more loudspeakers (excl. telephone sets, hearing aids and helmets with built-in headphones, whether or not incorporating a microphone); audio-frequency electric amplifiers; electric sound amplifier sets; parts thereof	
8519	Sound recording or sound reproducing apparatus	
8521	Video recording or reproducing apparatus, whether or not incorporating a video tuner (excl. video camera recorders)	
8522	Parts and accessories suitable for use solely or principally with sound reproducing and recording apparatus and with video equipment for recording and reproducing pictures and sound	
8523	Discs, tapes, solid-state non-volatile storage devices, "smart cards" and other media for the recording of sound or of other phenomena, whether or not recorded, incl. matrices and masters for the production of discs (excl. products of chapter 37)	
8525	Transmission apparatus for radio-broadcasting or television, whether or not incorporating reception apparatus or sound recording or reproducing apparatus; television cameras, digital cameras and video camera recorders	

Source: own elaboration

Overall, the data spanned the years 2010-2020 and was compiled at different levels of granularity: monthly, quarterly, bi-annual, and annual.[110] Overall, the initial training dataset included 31 612 observations (at the monthly level) and consisted of 246 unique combinations of producer country, reporting country, and product. However, due to significant numbers of missing values for some such combinations, the final dataset ultimately consisted of 26 488 observations and 203 unique combinations.

The final step of the process was creating a variable which served as a binary indicator of fraud, mentioned at the beginning of this section as necessary in the case of supervised training. The values of the indicator were determined through a thorough visual assessment of figures presenting the *import* and *export* time series for each country pair and product,[111] coupled with a comparison of the absolute and relative difference between import and export for each observation in order to improve accuracy. Table 23 provides a description of the observations classified as *fraudulent* and/or covered by the

[110] In the case of aggregated data, both the maximum and the minimum of the variable rcm_import were retained for each period.

[111] In order to reduce the risk of bias, the figures did not contain information on the country pair or product they represented.

domestic reverse charge mechanism and offers a comparison of observations classified as *fraudulent* and as *not fraudulent*.

Table 23: Observations classified as fraudulent or covered by the domestic reverse charge mechanism

	Number	Share of whole dataset	Categories
Observations with active Reverse Charge Mechanism	5 208	19.7%	16 categories, all considered high risk
Observations classified as fraudulent	1 169	4.4%	16 categories, all but one considered moderate or high risk, including six not covered by the RCM
Observations with active Reverser Charge Mechanism & classified as fraudulent	211	0.8%	six categories, all considered high risk

Source: own elaboration based on the training dataset derived from Intrastat.

Table 24: Comparison of fraudulent and non-fraudulent observations

	Fraudulent (*fraud_corr = 1*)	Not fraudulent (*fraud_corr = 0*)
Import value (EUR)	32 720 223	1 163 817 018
Export value (EUR)	77 982 710	1 224 330 819
Absolute difference between export and import (EUR)	45 262 487	60 513 801
Relative difference between export and import	0.58	0.05

Source: own elaboration based on the training dataset derived from Intrastat.

The figures resulting from this classification are presented below, in Figure 19 and Figure 20. The study team was guided by three assumptions: MTIC fraud would result in high absolute and relative differences between the reported values of import and export,[112] upticks in said differences would need to remain for at least three consecutive periods in order to minimise the risk of categorising outliers as fraud, and the difference between import and export could not remain stable over the analysed period – this would suggest that the discrepancies were likely caused by factors other than MTIC fraud (see Section V.a for an assessment of discrepancies in trade data). Although this experiment is to some degree reliant on the assumption that MTIC fraud is unlikely to occur when a domestic reverse charge mechanism for the relevant product is active, we chose to classify "suspicious" discrepancies as fraudulent even if a domestic reverse charge mechanism was in place, as classifying these instances as not fraudulent risked confusing the algorithm.

[112] Based on this, the study team classified absolute differences below EUR 1 mln as not fraudulent.

Figure 19: Examples of figures generated following the assessment and classified as fraudulent

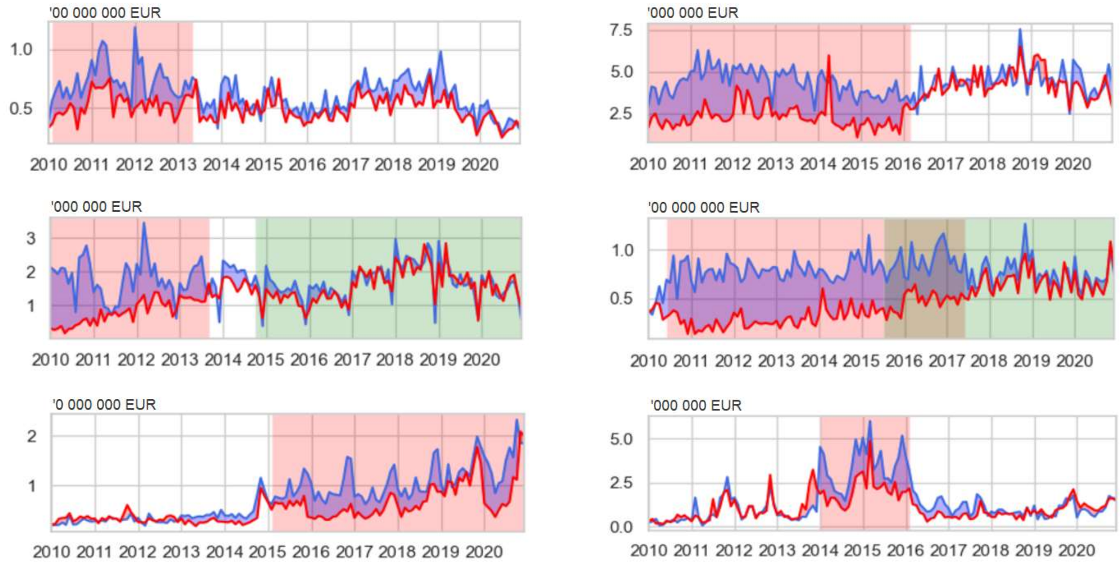
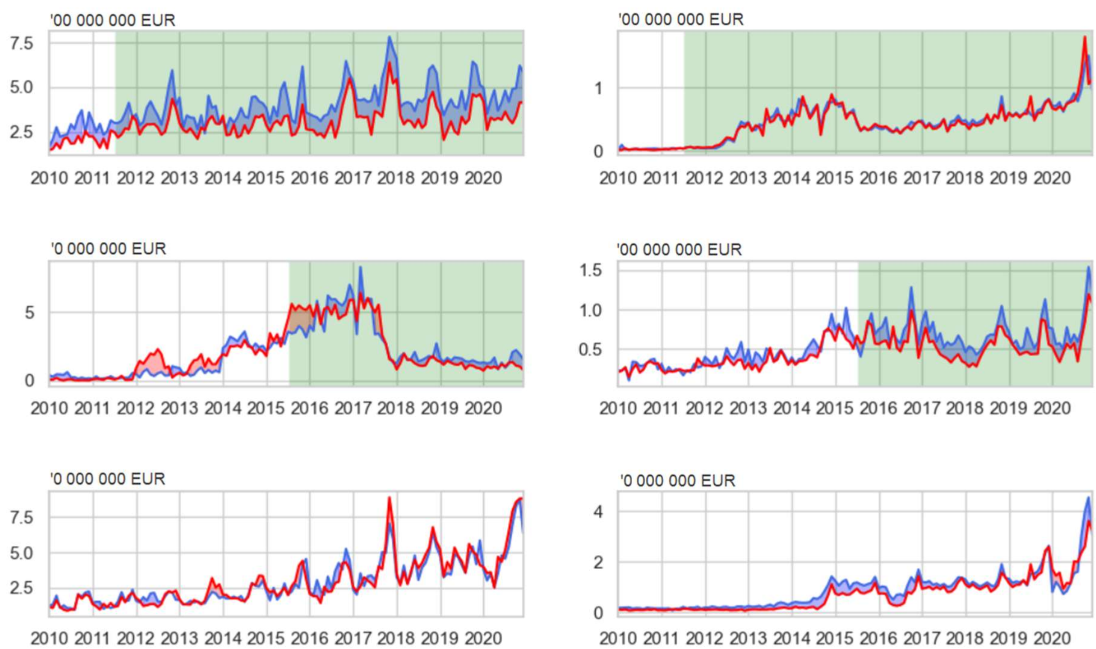


Figure 20: Examples of figures generated following the assessment and classified as not fraudulent



Source: own elaboration based on the training dataset derived from Intrastat.
Note: The y axis represents the trade volume (in EUR), the export series is blue and the import series is red. The area of the graph is coloured red for periods classified as fraudulent (fraud_corr = 1) and green for periods in which the domestic reverse charge mechanism for a given product and importing country was active (rcm_import = 1).

VII.c. Scenario #2

This section explores the application of three alternative classification methods for identifying fraudulent activities within the training dataset described in the preceding section and for estimating the scale of the MTIC gap. The primary objective of this phase of our study is to establish a robust starting point for calculating the scale of MTIC fraud within the EU context. Our focus is to effectively identify instances of potentially fraudulent activities within the training dataset, which will subsequently serve as a foundation for estimating the scale of the MTIC gap. The analysed models include econometric probit and logit models, and decision trees, specifically using the J48 algorithm.[113] The list of variables presented in the subsection below is used for both methods of classification.

Variable selection

To uncover the relationships between trade dynamics and the variable *fraud* we have formulated a set of explanatory variables that encapsulate various dimensions of trade behaviour. Our methodology involved analysing factors that influence trade, including export-import relationships, deviations from expected patterns, volatility, and standardized measures. By drawing from both empirical evidence and theoretical insights (see Chapter IV.a), we established a comprehensive set of explanatory variables that collectively offer a portrayal of trade dynamics (see Table 25).

Table 25: Variables defining trade dynamics

Variable	Description
expchange_t_t0	change of ICS (relative) with respect to the previous period (t-1)
expchange_t1_t	change of ICS (relative) with respect to the following period (t+1)
reexpchange_t_t0	change of re-export[114] (relative) with respect to the previous period (t-1)
reexpchange_t1_t	change of re-export (relative) with respect to the following period (t+1)
impchange_t_t0	change of ICA (relative) with respect to the previous period (t-1)
impchange_t1_t	change of ICA (relative) with respect to the previous period (t+1)
rel_deviation	relative deviation of ICS with respect to ICA at time (t)
rel_deviationt1	relative deviation of ICS with respect to ICA at time (t+1)
rel_deviationt0	relative deviation of ICS with respect to ICA at time (t-1)
averagedisp_20102018	average dispersion of ICS with respect to ICA over the entire period (2010-2018)
averagedisp	average dispersion of ICS with respect to ICA at time (t)
averagedisp1	average dispersion of ICS with respect to ICA at time (t+1)
averagedisp0	average dispersion of ICS with respect to ICA at time (t-1)

[113] See Box 4.

[114] Understood as Intra-Community Supply of the same category of goods to the origin of Intra-Community Acquisition in the same time period.

Variable	Description
stddevdisp	volatility (standard deviation) of relative dispersion across periods
export_std	volatility (standard deviation) of ICS
import_std	volatility (standard deviation) of ICA
trendexp	parameter/slope of the trend of ICS over the entire period
avg_relative_dispersion_linear	average relative dispersion of ICS from linear trend over the entire period
seasonal_factor	measure for the strength of seasonality
deviation_twoperiod	deviation of ICS value from the average ICA value from the two preceding periods
deviation_fourperiod	deviation of ICS value from the average ICA value from the four preceding periods
zerotrade	presence of zero trade (either ICS or ICA in any of the periods)
zscore_export_import	z-score for the difference between ICS and ICA
dtw	Dynamic Time Warping distance[115]

Source: own elaboration based on the training dataset derived from Intrastat.

Logit and probit models

Probit and logit models are widely used binary classification techniques that are particularly useful when predicting binary outcomes, such as, for example, fraud or no fraud. These models estimate the probability of a particular event occurring, making them well-suited for fraud detection, where the goal is to determine the likelihood of a transaction being fraudulent (see Box 4).

The study focused on different temporal granularities and predictor variables to create a precise yet parsimonious model capable of identifying fraudulent activities. The process employed could be split into the following steps:

1) An initial model was established using a comprehensive set of predictor variables. This model served as the foundation for subsequent analysis, ensuring a thorough consideration of potential indicators.

2) Temporal analysis was conducted across various timeframes, including monthly, yearly, quarterly, and biyearly intervals. This exploration aimed to uncover temporal patterns to determine the most informative granularity.

3) Collinearity and multicollinearity issues were addressed through correlation matrix analysis and Variance Inflation Factor (VIF) calculations. These procedures identified and mitigated collinear relationships among predictor variables. The correlation matrix identifies pairs of variables that are highly correlated, allowing for the removal of unwanted predictors. VIF quantifies how much multicollinearity affects the variance of regression coefficients. Thus, high VIF values indicate the need to remove or adjust collinear variables.

4) Comparison of the probit and logit models was carried out for each temporal granularity. To optimize the model's performance and identify the most influential variables, a systematic process

[115] See discussion of this vaiable in Section VII.d.

of variable sel ection was employed. One by one, predictor variables were tested by introducing them into the model and evaluating their impact on model fit and predictive accuracy. Variables that demonstrated a significant improvement in model fit were retained, while variables with minimal impact or potential issues of overfitting were systematically excluded. This stepwise approach ensured that only the most relevant and non-redundant variables were included in the final model.

5) After rigorous evaluation, the logit model measure by the accuracy of classification consistently demonstrated superior performance across temporal scales.

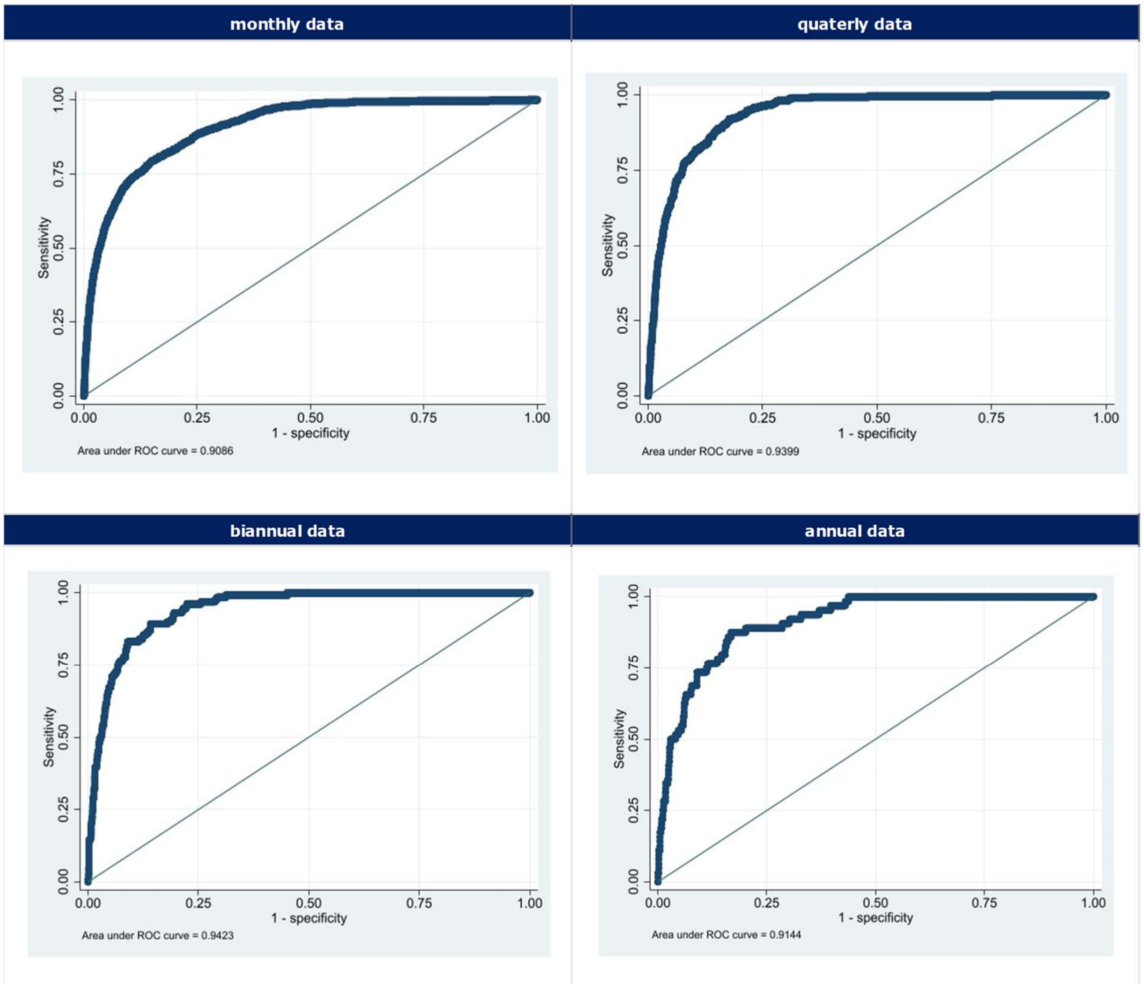
Model selection

Our evaluation of the predictive models involved assessing their performance based on specific criteria, providing a comprehensive perspective on their effectiveness. As a first step, we investigated omitted variables due to collinearity concerns. In the monthly and quarterly analyses, variables such as *reldisp_20102018*, *averagedisp*, *ave~20102018*, *deviationtwoperiod*, *averagedisp0*, and *zero_export_import* were excluded. The biyearly and annual models followed a similar trend, omitting *reldisp_20102018*, *averagedisp* and *zero_export_import* due to the same concerns.

As a next step, we examined the discriminatory capacity of the models using the area under the localization receiver operating characteristic (LROC) curve, which is a graphical representation of a model's ability to predict rare events – in this case instances of fraud – by capturing the relationship between the cumulative proportion of positive outcomes and the cumulative proportion of false alarms. In essence, the LROC curve helps us evaluate the trade-off between sensitivity (true positive rate) and specificity (1 - false positive rate) in detecting potential fraudulent transactions.

Across the temporal perspectives, the monthly model achieved an LROC value of 0.9086, indicating a strong predictive ability. Similarly, the quarterly model yielded a high Area Under the Receiver Operating Characteristic Curve (AUC-ROC) value of 0.9399, demonstrating its effectiveness in distinguishing fraudulent activities. Notably, the biannual model exhibited the highest LROC value of 0.9423, signifying its robust discriminatory power. The annual model, while slightly lower at 0.9144, still showcased a notable capacity for differentiation. For each temporal perspective (annual, biannual, quarterly, and monthly), we provide LROC curves that visually depict the performance of our predictive models in identifying instances of fraud (Table 26).

Table 26: LROC curves across different time periods



Source: own elaboration based on the training dataset derived from Intrastat.
Note: the higher the placement of the ROC curve, the greater the explanatory power of the model.

Lastly, we assessed the explanatory power of the models using the Pseudo R-squared value. The biyearly model displayed the highest Pseudo R-squared value of 0.4018, indicating its ability to explain a significant portion of the variation. The annual model followed suit, with a Pseudo R-squared value of 0.3472, suggesting a moderate yet meaningful level of explanatory capacity. The outcome suggests that the biyearly perspective provides the most balanced view of trade dynamics, making it the preferred approach for identifying potentially fraudulent activities in our dataset.

Table 25 presents the classification results of the model. The model's predictions are categorized into positive (fraudulent) and negative (non-fraudulent) classes (D and ~D) based on their alignment with the target class.

Additionally, the table includes two performance metrics:

- Sensitivity, which gauges the model's ability to correctly identify positive instances and is computed by dividing the number of correctly classified "D" instances by the total actual instances of class "D." In this case, sensitivity is approximately 55.95%.

- Specificity, which assesses the model's accuracy in identifying negative instances and is calculated by dividing the number of correctly classified "~D" instances by the total actual instances of class "~D." The specificity rate is around 96.01%.

The results of the logit regression are provided in Table 26.

Table 27: Classification results of the logit model

	True		
Classified	**D**	**~D**	**Total**
+	47	130	177
-	37	3 127	3 164
Total	84	3 257	3 341
Sensitivity			0.5595
Specificity			0.9601

Source: own elaboration based on the training dataset derived from Intrastat.

Table 28: Logit regression coefficients

Variable	Coefficient
avg_export_newdate	0.000
expchange_t_t0	-1.557
expchange_t1_t	-0.018
reexpchange_t_t0	-0.162
reexpchange_t1_t	-0.003
rel_deviation	-0.035
rel_deviation1	0.024
rel_deviation0	0.075
averagedisp_20102018	-41.548
averagedisp1	0.848
averagedisp0	1.002
stddevdisp	-0.079
stddevexp	-2.669
stddevimp	-0.621
trendexp	0.000
trend_coeff	0.000
relative_dispersion_linear	-0.066
avg_relative_dispersion_linear	0.278
seasonal_factor	0.692
deviation_twoperiod	0.142
deviation_fourperiod	-1.113
zscore_export_import	0.821
export_standardized	-0.068
import_standardized	0.060

Variable	Coefficient
dtw	-0.162
_cons	-4.151

Source: own elaboration based on the training dataset derived from Intrastat.

Simulation

We applied our logistic regression model developed on a training dataset to a larger and more comprehensive dataset. Our goal was to capitalize on the model's proven predictive capabilities to identify potential instances of fraud across a broader range of trade scenarios. Our approach involved deploying the established coefficients to the new dataset, thereby allowing automated fraud detection. The coefficients, derived from the initial training dataset, encapsulate the relationships between the explanatory variables and the dependent variable, *fraud*. Our assumption was that these relationships remain consistent across the larger dataset and that the model was well-trained on the smaller dataset, and thus, the coefficients can be utilized as-is to calculate the log-odds of each observation being indicative of fraud.

The process of applying the model to the larger dataset entailed straightforward calculations. For each observation in the new dataset, we utilized the learned coefficients to compute the log-odds of *fraud* being 1. Subsequently, we transformed these log-odds into probabilities using the logistic function. This automated approach allowed us to systematically assess the likelihood of fraudulent behaviour for each transaction in the larger dataset.

Out of all observations, 44,293 were classified as fraudulent, representing approximately 6% of all observations. The Netherlands had the highest absolute number of fraud cases detected, with 10 817 instances out of a total of 95 595 transactions (11.3%) (see Table 29).

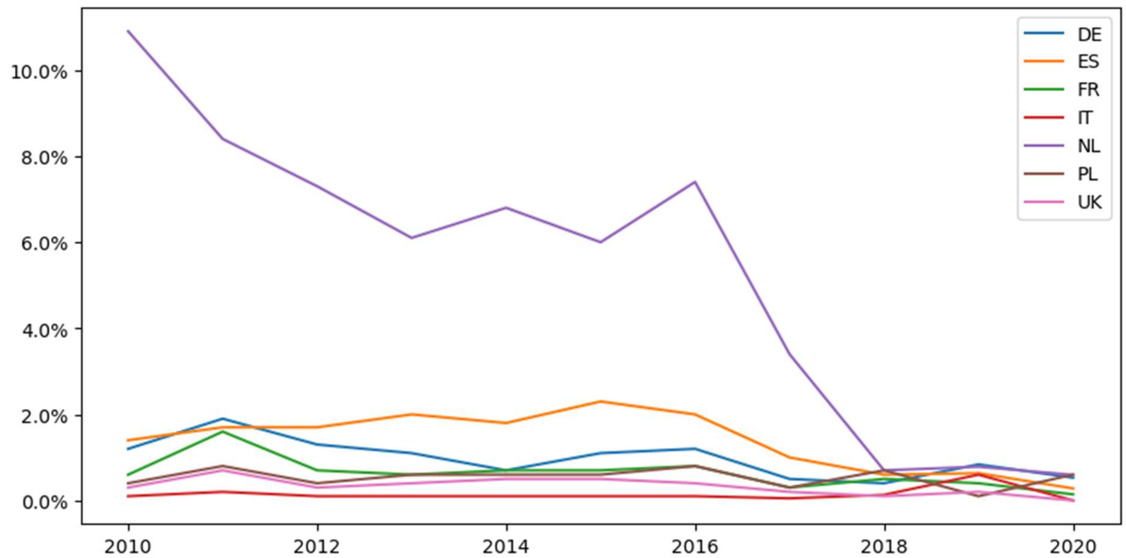
Table 29: Number of fraudulent observations by country

	Fraud	Total	fraud (%)
DE	9 921	120 838	8.2%
ES	5 082	95 278	5.3%
FR	6 524	114 321	5.7%
GB	4 459	91 448	4.9%
IT	3 475	106 488	3.3%
NL	10 817	95 595	11.3%
PL	4 015	78 992	5.1%

Source: own elaboration based on the training dataset derived from Intrastat.

Drawing upon the outcomes of our model's classification, we proceeded to conduct an estimation of the cumulative value associated with fraudulent transactions. This quantification was achieved through a summation of the dispersions existing between mirror statistics exclusively for transactions flagged as fraudulent by our predictive model. The sum was then multiplied by the standard statutory VAT rate to estimate the potential VAT revenues lost due to these discrepancies. This rate was selected as a pragmatic compromise, seeking to capture a middle-ground approximation of the potential impact, acknowledging that variations in VAT rates, trade volumes, and other factors could influence the final value. Through this methodology, we arrived at an estimated mean yearly value of EUR 9.48 billion of revenue potentially lost yearly to MTIC fraud in goods traded between seven Member States in our dataset between 2010 and 2020.

Figure 21: Share of estimated MTIC fraud VAT in the total VAT compliance gap (note: MTIC losses are calculated on a sample of country partners)



Source: Own elaboration based on the training dataset derived from Intrastat.

Decision tree

As discussed in Box 4, one of the main types of classification algorithms are decision trees, with C4.5 being one of the most popular representatives of those algorithms. This algorithm has been tried and tested by various researchers over the years and should be considered a good starting point for this exercise. It is important to note that C4.5 represent a non-parametric class of algorithms which in general perform significantly better when provided with a large training dataset, are relatively slow and are prone to over-fitting. Another limitation is interpretability – as long as the tree contains a small number of nodes, it can provide a pretty straightforward, intuitive picture, but with more complex trees the interpretation becomes much less clear.

A practical implementation of the C4.5 algorithm is offered by a free data mining tool Weka (the specific implementation is called J48). Among the significant advantages C4.5 algorithm are 1) its ability to handle training data with some missing values 2) ability to handle continuous attributes, not just discrete (although this was not taken advantage of) and 3) its ability to apply pruning, which limits the risk of over-fitting. This last issue – over-fitting – can still occur if the tree is allowed to create too many branches. This can be controlled with pruning parameters – mainly the minimum number of instances in each node and confidence factor (the level of accepted error rate). Different specifications of decision tree were tested on the training dataset described in Section X.b – all specifications were calculated using a monthly dataset where variables were discretized in advance. The best performing decision tree is presented in Figure 24, with decision rules and nodes represented in a flowchart. Table 30 presents the confusion matrix[116] for the same model, while Figure 23 shows the ROC curve – separately for

[116] The confusion matrix is a simple representation of counts of predicted and actual values. The observations are divided into four groups – True Negative, False Negative, False Positive and True Positive.

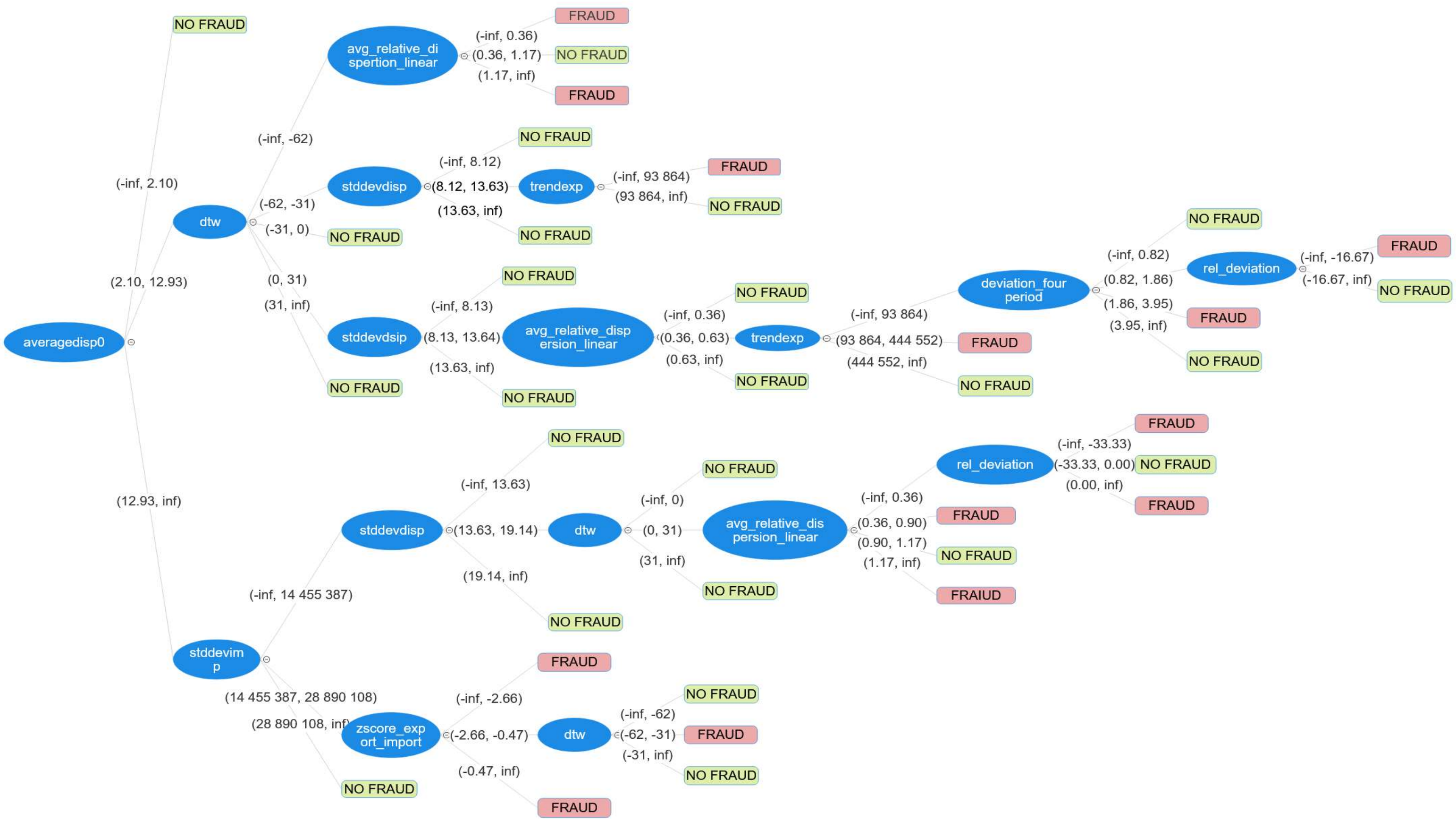
positive and negative labels. It is important to note that, in the case of this particular training dataset, the number of positive labels is very small compared to the negative ones – thus, even if the overall accuracy of prediction is very high with 98.4% of labels predicted correctly, this is not very informative. Much more telling of the true ability of the decision tree is the share of correctly classified positive labels – in this case it is 45%. It is also worth to note that the decision tree is very rarely falsely classifying negative labels, which suggests that those cases are pretty clear cut and there is a good potential for improvement.

Table 30: Confusion matrix for best performing C4.5 decision tree

		Actual	
		0	1
Predicted	0	25 792	372
	1	25	299

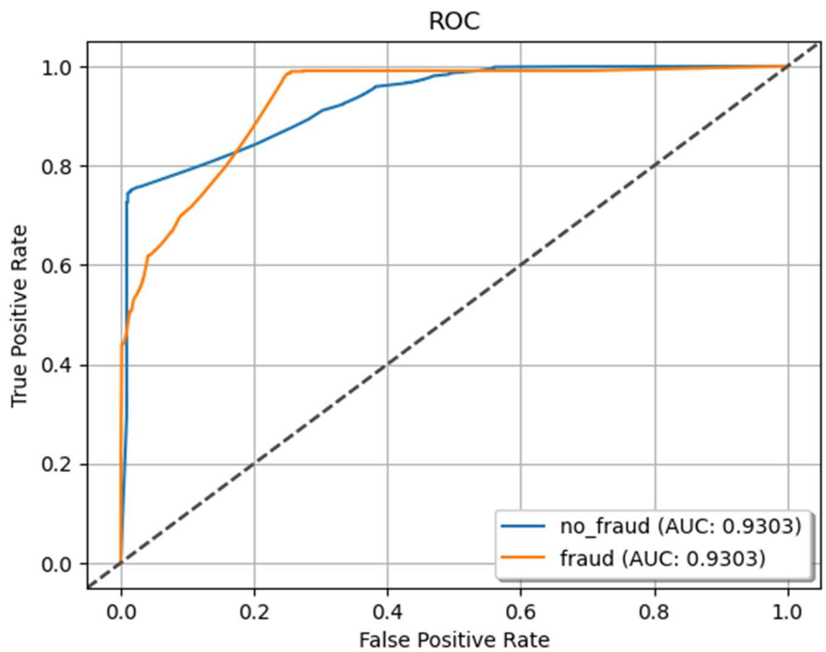
Source: Own elaboration.

Figure 22: Flowchart for best performing C4.5 decision tree, lines indicate decision rules while boxes represent nodes



Source: Own elaboration.

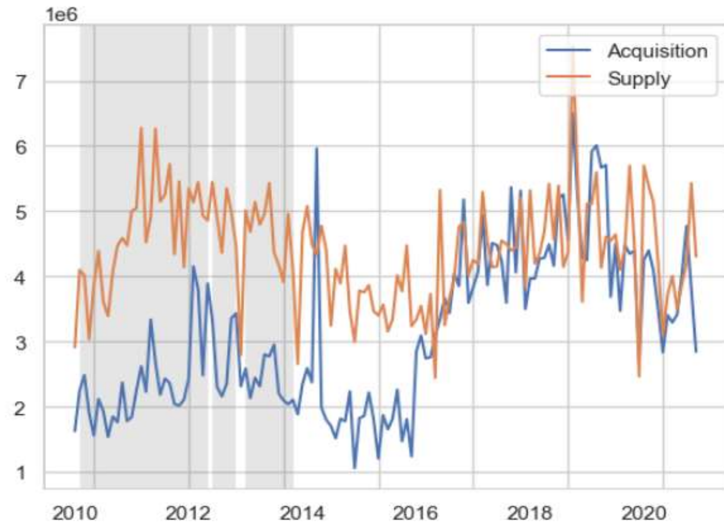
Figure 23: ROC curve for the best performing C4.5 decision tree



Source: Own elaboration

To better visualize the idea behind decision tree classification and its usefulness for the estimation of the MTIC gap, the example of one matched time series for Acquisition and Supply is presented in Figure 24. This time, in contrast to Figure 19 and Figure 20 (with labels created by the team), the grey areas are periods where the classification tree predicted the positive fraud label. In this particular case the decision tree correctly classified some periods where reported Intra-Community Supply exceeds reported Intra-Community Acquisition in a pattern which was considered to be related to MTIC fraud (temporary structural break in the mirror statistics). On the other hand, it seems that the decision tree incorrectly classified periods around 2015 and 2016, assigning them a negative fraud flag – we would expect those periods to be shaded in grey as well, as they seem to follow a similar pattern to the 2010-2014 period. This shows that there is still room for improvement, whether in the choice of method, the specification of decision tree or choice of variables.

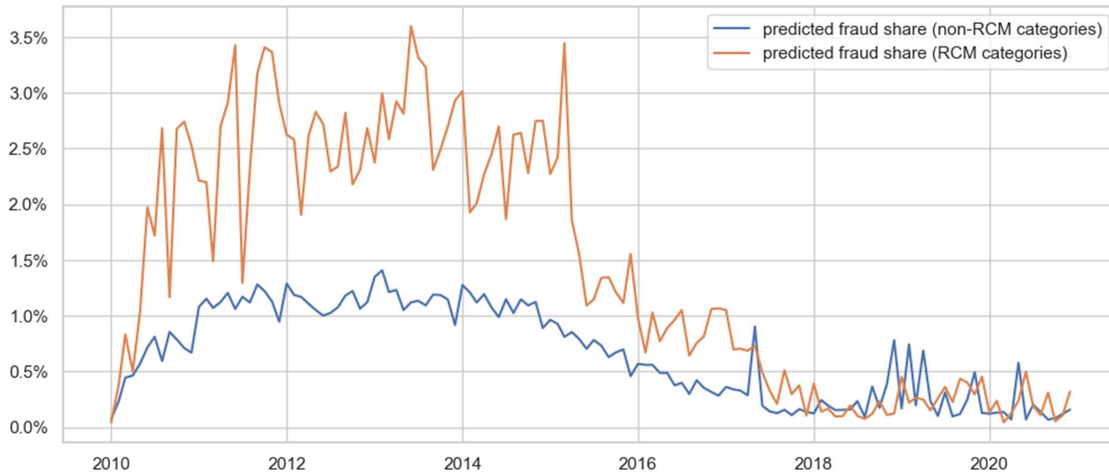
Figure 24: Example of predicted classification of fraud label based on a C4.5 decision tree



Source: Own elaboration.
Note: The grey shading denotes the period classified as "fraud".

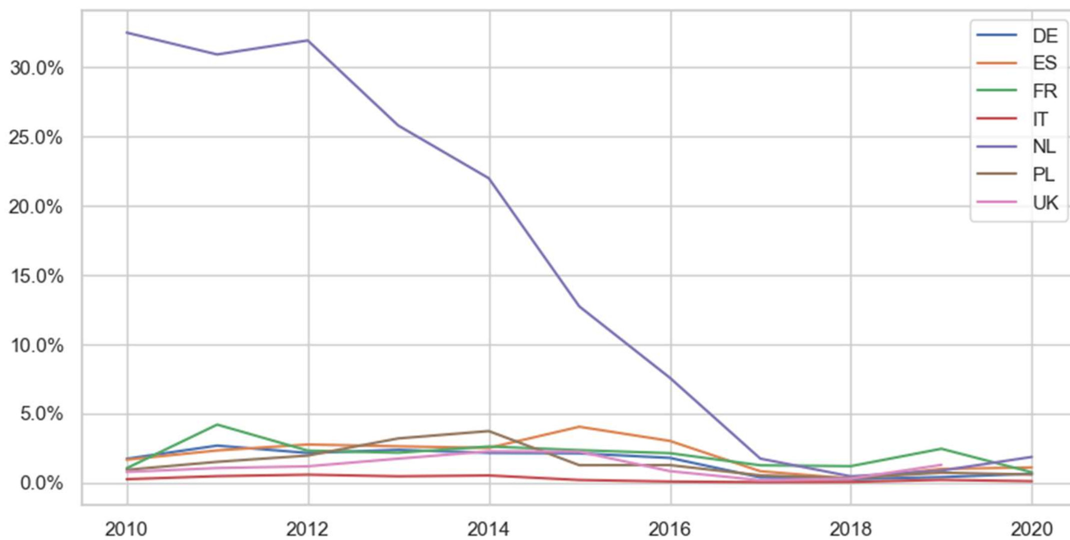
In the next step, the rules produced by the decision tree were applied to the largest of the datasets which was described in Chapter VIII (7 countries (42 pairs), 1 227 CN4 product categories) in order to get a better understanding of the estimate of the scale of MTIC fraud. For this purpose, we constructed a simple variable which calculates the excess of Supply over Acquisition in periods which were classified as fraud by the decision tree (in the above-mentioned example this can be visualized as areas between Supply and Acquisition in periods marked grey). The relative value of that excess to the value of the Supply is presented in Figure 24. This ratio was presented for two separate groups – CN4 products which were never under the domestic reverse charge mechanism (blue line) and categories which were under the domestic reverse charge mechanism at some point in time (orange line). Interestingly, there seems to be a large difference between those two groups - categories in which the domestic reverse charge mechanism was introduced show a significantly larger share of supply which is labelled as fraud. This, to some extent, could confirm that those categories were the focus of the fraudsters at some point, were correctly identified by the authorities as such and thus the domestic reverse charge mechanism was introduced. A similar conclusion can be drawn from the large drop around 2015 in share of supply identified as fraud - during this time a large number of countries within the sample introduced their own legislature on Reverse Charge Mechanism (see Table 22). Preliminary estimates of MTIC fraud scale using decision trees suggest that those actions were successful. A more detailed analysis is necessary in order to identify whether, in response to those actions, the fraudsters switched to different products. It is very important to remember that the sample which was used to arrive at those numbers is just a fraction of the full Intrastat dataset – each country has just six trade partners, instead of 26. Based on this subsample the value of MTIC fraud between 2010 and 2020 was estimated to be around EUR 9.4 billion annually, i.e., nearly the same value as pointed by the logit model. The relative share potential related VAT losses (assuming 20% tax rate) in total VAT compliance gap was presented in Figure 26.

Figure 25: Share of excess Supply in observations classified with the positive fraud label in total Supply, separately for CN4 categories under Reverse Charge Mechanism (orange line) and other (blue line)



Source: Own elaboration.

Figure 26: Share of estimated MTIC fraud VAT losses in the total VAT compliance gap



Source: Own elaboration.
Note: MTIC losses are calculated on a sample of country partners.

Table 31: Share of estimated MTIC fraud in total supply, by origin and destination country (in total, 2010-2020)

		Destination						
		DE	ES	FR	UK	IT	NL	PL
Origin	DE	.	0.6%	0.8%	0.6%	0.2%	2.9%	0.8%
	ES	0.5%	.	0.8%	0.3%	0.5%	0.6%	0.5%
	FR	0.3%	0.3%	.	0.7%	0.6%	1.7%	0.9%
	UK	1.0%	0.6%	0.9%	.	0.8%	1.1%	0.4%
	IT	0.5%	1.2%	1.0%	1.2%	.	1.6%	0.5%
	NL	0.4%	0.1%	0.2%	0.4%	0.1%	.	0.3%
	PL	0.6%	0.4%	0.4%	0.3%	0.1%	0.7%	.

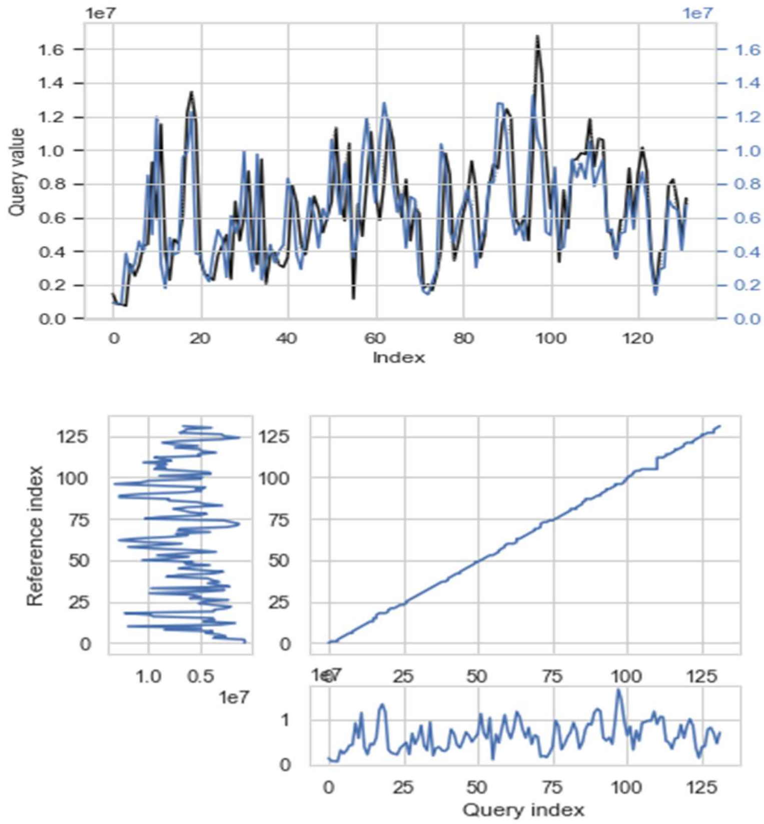
Source: Own elaboration.

VII.d. Scenario #3

As previously mentioned, one of the standard approaches to unsupervised classification is k-means clustering (for more details on k-means see Box 3). The simplicity of that approach is its undeniable strength, but its basic version is not well suited to time series applications. In recent years new approaches specifically designed for time series analysis have been proposed - one such algorithm is k-shape (Gravano, 2016). Similarly to k-means, it is based on an iterative refinement procedure which assigns observations (in this case time series) to clusters, adjusts clusters' centroid positions to better fit the assigned observations and repeats the process. The main difference of that approach compared to k-means is that it uses shape-based distance measure instead of simple Euclidean distance.

The main issue is that the simple measure of discrepancy produced by subtracting Acquisition from Supply contains a lot of the same noise which is observed in the original time series. As mentioned in Chapter VIII, there are some known issues within Intrastat dataset such as seasonal hikes in discrepancies of mirror statistics or lags which can obscure the overall patterns in the dataset and make it hard to distinguish between discrepancies associated with MTIC fraud and others. The perfect indicator of mirror statistic discrepancy size should consider not only the relative size of the gap between reported values of trade, but also slopes of Acquisition and Supply, long-term patterns of each, and the time lag which might be present between those two series (we can think of those properties as shapes of two time series, which can be scaled, stretched or shifted in time). One of the promising candidates for producing a measure which, to some extent, takes those considerations into account comes from the dynamic time warping algorithm, which can be used in partial shape matching between time series. This algorithm is primarily used to measure the similarity between two time series, which can differ in speed, length, scale etc. (one of the most famous applications of this algorithm is voice recognition). Figure 27 below shows the algorithm in action (on the same example as presented in Chapter VIII, Figure 13). Although the shapes of Supply (black line) and Acquisition (blue line) variables are mismatched through most of the observed period (and subtracting one from the other would produce a variable suggesting that there is significant discrepancy between reported values), the dynamic time warping algorithm matches the shapes of time series and produces an indicator which follows a path close to a diagonal line (see the bottom half of the Figure 27).

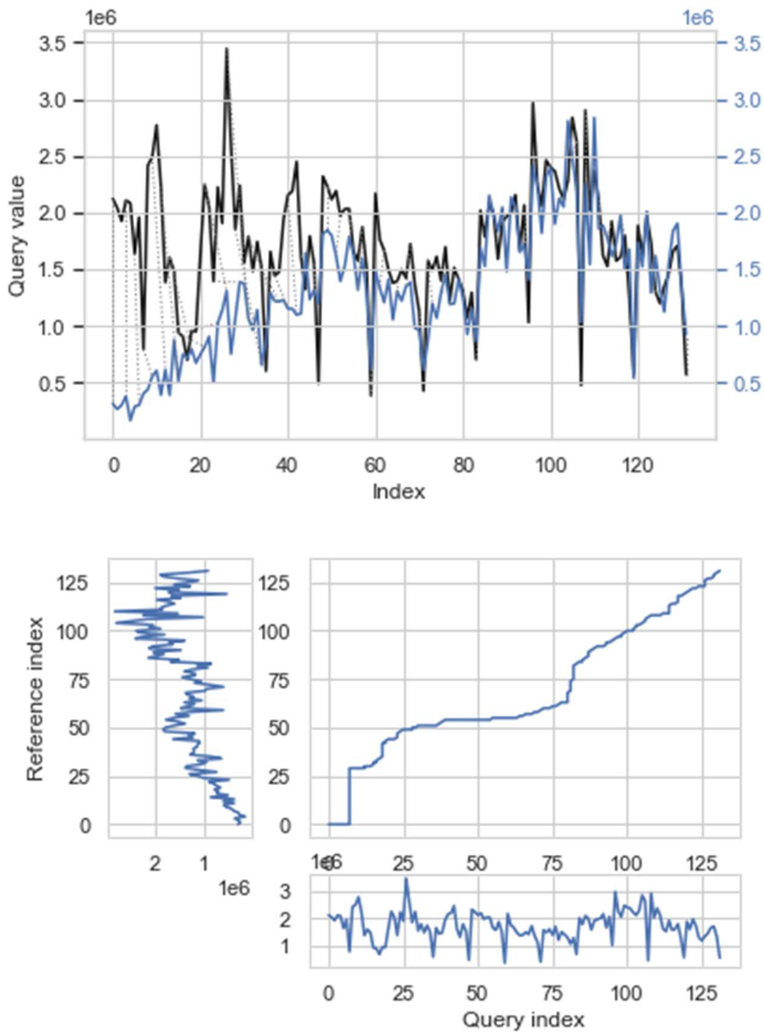
Figure 27: Example of dynamic time warping matching on two time series with a one period time lag



Source: Own elaboration.

Another example (Figure 28) presents a different situation - here two time series are clearly mismatched through the first half of the observed period and then converge after month 60 (and after month 80 the values become almost identical). The product of matching presented in the bottom part of the figure below tells a similar story – the value of the match index is far from the diagonal line through the first part, showing that the algorithm struggles to match the values within that period, but at the later stage the line becomes nearly diagonal, suggesting good alignment of the two time series. The indicator produced in such a way helps to filter out much of the noise observed in the original data and provides simplified characteristics of anomalies.

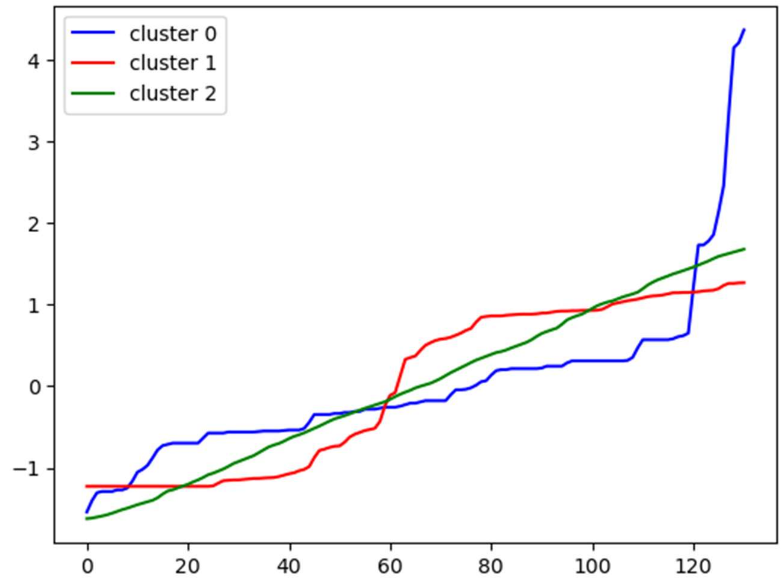
Figure 28: Example of dynamic time warping matching on two time series with a temporary structural break



Source: Own elaboration.

Figure 29 presents the effect of k-shape clustering conducted on the sample of mirror statistic discrepancies calculated with dynamic time warping. For the sake of simplicity, this illustrative example was prepared on a relatively small sample containing just 39 time series, clustered into three groups. In practice, a significantly larger number of clusters would be necessary to account for the various ways in which Intra-Community Supply and Intra-Community Acquisition variables misalign (e.g., a discrepancy which lasts for 20 months starting at the beginning of the analysed period produces a different shape than a discrepancy lasting for 6 months which occurred at the end of that period). In this particular example the green line represents all cases where the Supply and Acquisition time series are closely aligned, and the red and blue lines represent certain types of misalignments. It is important to note that, although the method itself is unsupervised, the interpretation of each cluster shape is a separate task and has to be done manually. On top of that, the classification is prepared at the level of the whole time series rather than of a single observation, which means that another step is required in order to identify periods in which MTIC fraud was suspected. Because of this, at this stage the method was not used to estimate the nominal scale of MTIC fraud.

Figure 29: Centroid shapes produced with k-shape clustering (3 clusters) on the sample of 39 time series constructed with dynamic time warping



Source: Own elaboration.

VII.e. Summary

The experimental partial implementation of two methodological scenarios based on Intrastat trade statistics and the combination of econometric and machine learning techniques confirm the observations made in the assessment and comparison with other approaches. Both classification and clustering techniques are promising tools to filter that patterns in the data that are expected indicate MTIC fraud. Various measures quoted in this chapter including sensitivity and specificity ratio, R-squared and AUC confirm this observation. Similarly, the estimated results including the share of fraudulent transactions and the value of MTIC fraud are largely consistent with the expectations regarding their scale and change over time. At the same time, not all the issues observed could be solved and the generated are measures of fit and simulation results raise some questions that could not be answered during this test.

The full-fledged implementation will require more thorough analysis in making multiple and difficult choices involved in the modelling including the choice of the model, variables, filtering, parameters of the model, granularity of the dataset and many more. The full analysis will also require the implementation of a modelling algorithm using different tools and able to handle over 15-times larger dataset. Already, at the stage of this experimental implementation, the study team had to group some calculations in blocks to meet the limit of software limitations.

The analysis showed that all involved algorithms, econometric probit and logit model, decision trees and k-shape algorithm, are suitable and promising analytical method for Phase II of the study. Data mining techniques appeared to be very effective, yet the rules that they produce, and the results obtained are more difficult to interpret than in the case of the econometric models. This suggests that the full implementation of the estimation should consider alternative techniques – both classification and clustering algorithms and their different sub-types. Parallel implementation of different algorithms up to the moment, when some specific approaches prove to be superior, will allow cross validating and decreasing uncertainty around the final results.

VIII. Conclusion and recommendation

Although tax administrations have amassed significant experience in detecting MTIC fraud and the academic literature has already explored a number of variables and methods used to track anomalies present in economic variables and revenue data, there is no simple one-fits-all solution to estimating the scale of MTIC fraud in all EU Member States using a standardised methodology.

The methodologies for estimating the MTIC gap, as in the case of other tax gaps, could traditionally be grouped into top-down approaches, based on aggregate macroeconomic and trade statistics, and bottom-up approaches based on VAT reporting and audit results. Out of the two method groups, bottom-up methods appear to be characterised by the highest precision. On the one hand, the approach based on transaction-level data from VIES and VAT returns would allow tracking fraud for each individual in the entire population of taxpayers without posing significant measurement problems. On the other, the method based on audit results would allow to use measurement of the scale of individual fraud and extrapolate it to the entire tax base and all taxpayers. Yet, despite these advantages, bottom-up methods based on individual-level administrative data carry many limitations which significantly impact the feasibility of their widespread use. Chief among them is the issue of data availability and the high costs associated with gathering the data needed for producing the estimates.

Although the above-mentioned methods might be considered superior when assessing the precision and granularity of the results alone, in the context of this study only a small fraction of Member States would be able to share the data necessary to perform such estimations, largely due to confidentiality considerations. Nevertheless, it may be possible to circumvent this issue by allowing for such estimations to be conducted by Member State administrations using a standardised methodology, the resulting estimates would then be gathered by the European Commission or its contractor in a single publication. Another issue faced when dealing with methods reliant on audit data is the resource intensiveness of audits, which translates to these methods being by far the most resource-intensive to implement. Although methods comparing data in VIES and from VAT returns are available for all administrations, they still remain relatively costly, cover only the period from 2020 and would measure only the broader cross-border fraud, without distinguishing MTIC fraud in particular.

In general, the methodologies using macro-level figures promise a lower degree of accuracy. Accurate estimation using macro-level figures requires controlling for noise in the underlying figures, which may not be possible even with the most sophisticated numerical tools. The methods based on macro indicators may create a black-box effect – the inability to understand the drivers and sources of the estimated fraud and its variation in time. In addition, depending on the specific method selected, the underlying datasets carry their own unique limitations, such as weak coverage of fraud in services or problems with distinguishing between different types of fraud schemes. The most promising method based on publicly available data uses trade data, which are available for narrow categories of products, thus avoiding some of the problems related to observing broader indicators.

The assessment, conducted using a formalised framework reflecting the multiple objectives of the EU MTIC gap calculation, has shown that the methodological scenario based on Intrastat statistics is the most promising. Combining the use of data mining techniques, discrepancies between mirror registers and granular data on the magnitude of trade and its shift in time offers a methodological scenario that could be implemented for all Members States at a sufficient level of accuracy and would meet all the criteria set out for the study to the highest degree. The second-best methodological scenario, when considering the feasibility of the MTIC gap estimation being conducted by the study team, is the method based on VIES and VAT returns. Although this method appears to be superior in terms of precision, the limited access to this data by the study team diminishes its operability and feasibility.

As there are some uncertainties around the assessment of approaches that have not been implemented, the study team decided for the experimental partial implementation of the two methodological scenarios based on Intrastat trade statistics and the combination of econometric and machine learning techniques. This analysis has demonstrated that both classification and clustering techniques are promising tools capable of detecting irregularities that could be attributed to fraud. Various measures, including the sensitivity and specificity ratio, R-squared and AUC, confirmed this observation. Similarly, the estimated results, which include the share of fraudulent transactions and the value of MTIC fraud are largely consistent with expectations regarding their scale and change over time. At the same time, not all of the issues observed could be solved and the generated measures of fit and simulation results raise some questions which could not be answered during this test.

Based on these takeaways, it is recommended for the European Commission to continue the development of the methodology to estimate forgone VAT revenue due to MTIC fraud in the EU and EU Member States under Phase II of this study. In particular, it is recommended that the methodological scenario based on Intrastat data and classification data mining techniques (Scenario #2) be used for this purpose. Phase II should extend the implementation of the methodology to all Member States and carefully reconsider all the modelling and data compilation decisions that must be made in this process. It is also proposed to further explore the availability of data from VAT returns and VIES and, should this data become available, carry out an experimental implementation of the methodological scenario based on matching these figures (Scenario #5) for selected Member States. Such an approach would allow proceeding with the best possible feasible approach, validate it and make a fully informed assessment of the two competing methodological scenarios.

Bibliography

Agha, A., & Haughton, J. (1996). Designing VAT systems: Some efficiency considerations. *The Review of Economics and Statistics, 303-308*.

Ainsworth, R., et al. (2018). A VAT Coin Solution to MTIC Fraud: Past Efforts, Present Technology, And the EU's 2017 Proposal. *Scholarly Commons at Boston University School of Law. Boston University School of Law*.

Ainsworth, R. T. (2010). MTIC (VAT FRAUD) in VoIP-Market Size $3.3 b., *Boston University School of Law Working Paper No. 10-03*.

Ainsworth, R. T. (2011). Technology Solves MTIC-VLN, RTvat, D-VAT Certification. *Boston Univ. School of Law Working Paper, (11-07)*.

Alexopoulos, A., Dellaportas, P., Gyoshev, S., Kotsogiannis, C., & Pavkov, T. (2020). *Detecting network anomalies in the Value Added Taxes (VAT) system. A TARC Policy Analysis Report*.

Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed.). Cambridge, MA: MIT Press.

Arango, C. A., Misas, M., & López, E. (2006). Economía subterránea en Colombia 1976-2003: una medición a partir de la demanda de efectivo. *ENSAYOS SOBRE POLÍTICA ECONÓMICA*.

Bishop, C. M. (2006). *Pattern recognition and machine learning. Information science and statistics*. New York: Springer.

Baesens, B., Vlasselaer, V., & Verbeke, W. (2015). *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. A Guide to Data Science for Fraud Detection*. John Wiley & Sons.

Barbone, L., Bird, R. M., & Vázquez Caro, J. (2012). *The costs of VAT: A review of the literature.* CASE Network Reports, (106).

Borges, H., Akbarinia, R., Messeglia, F. (2021). *Anomaly Detection in Time Series*. Transactions on Large-Scale Data-and Knowledge-Centered Systems L (46-62), Springer Berlin Heidelberg

Borselli, F. (2011). Organised VAT fraud: features, magnitude, policy perspectives. *Policy Perspectives (October 31, 2011). Bank of Italy Occasional Paper, (106)*.

Borselli, F., Fedeli, S., & Giuriato, L. (2015). Digital VAT carousel fraud: A new boundary for criminality? *Tax Notes International, 77(8)*, 707.

Braml, M. T., & Felbermayr, G. J. (2021). The EU self-surplus puzzle: an indication of VAT fraud? *International Tax and Public Finance, 29(5)*.

Breusch, T. (2005). Estimating the Underground Economy using MIMIC Models. *Econometrics,* University Library of Munich, Germany.

Buettner, T., & Tassi, A. (2023). VAT fraud and reverse charge: empirical evidence from VAT return data. *International Tax and Public Finance*.

Carrère, C., & Grigoriou, C. (2014). Can Mirror Data Help To Capture Informal International Trade? *UNCTAD Blue Series Papers 65, United Nations Conference on Trade and Development*.

CASE. (2015). *Ocena czy poprzez wprowadzenie od dnia 1 października 2013 r. odwróconego obciążenia VAT na niektóre wyroby stalowe ograniczono nadużycia w rozliczaniu podatku VAT w obrocie tymi towarami oraz czy warunki funkcjonowania przedsiębiorstw w branży objętej zmianami uległy poprawie.* Ministerstwo Finansów.

Cejková, E., & Zídková, H. (2019). How To Estimate The Size Of Carousel Fraud?. In *Proceedings of FEB Zagreb International Odyssey Conference on Economics and Business* (Vol. 1, No. 1, pp. 385-398). University of Zagreb, Faculty of Economics and Business.

Cerioli, A., Barabesi, L., Cerasa, A., Menegatti, M., & Perrotta, D. (2019). Newcomb–Benford law and the detection of frauds in international trade. *Proceedings of the National Academy of Sciences, 116*(1), 106-115.

Chandola V., Benerjee A., Kumar V. (2009). Anomaly Detection: A Survey. *CM Computing Surveys, 1-72*.

Chow, Sy-Miin, Ho, Moon-Ho, Hamaker, Ellen, & Dolan, Conor. (2010). Equivalences and differences between structural equation and state-space modeling frameworks. *Structural Equation Modeling, 17*.

Christie, E., & Holzner, M. (2006). *What explains tax evasion? An empirical assessment based on European data* (No. 40). WIIW Working Paper.

David, P., & Semerád, P. (2014). Possibilities of measuring tax evasion related to fuel sale. *Procedia Economics and Finance, 12*, 121-129.

DG TAXUD (2018). *The concept of Tax Gaps. Report III: MTIC Fraud Gap estimation methodologies.*

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., & Vapnik, V. N. (1997). Support vector regression machines. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9* (pp. 155-161). Cambridge, MA: MIT Press.

Dybka, P., Kowalczuk, M., Olesiński, B. et al. (2019). Currency demand and MIMIC models: towards a structured hybrid method of measuring the shadow economy. *International Tax and Public Finance, 26*(1), 4–40.

Eskander, A. (2017). Recent measures countering MTIC fraud. *Taxpro Journal*.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226-231). Portland, OR: AAAI Press.

European Commission (2017). *Study contributing to an Impact Assessment on Council Directive 2008/118/EC concerning the general arrangements for excise duty Final Report Volume 1 – Main Text*, ISBN: 978-92-79-72115-1.

European Commission, CASE, Poniatowski, G., Bonch-Osmolovskiy, M., Śmietanka, A., Pechcińska. (2021). *Study and reports on the VAT Gap in the EU28 Member States. 2021 Final Report*, Publications Office of the European Union, Luxembourg. Retrieved from [link]

European Commission, Directorate-General for Taxation and Customs Union, Poniatowski, G., Bonch-Osmolovskiy, M., Śmietanka, A., et al. (2022*). VAT gap in the EU: report 2022*, Publications Office of the European Union.

European Court of Auditors. (2015). *Tackling intra-Community VAT fraud: More action needed* [PDF File]. Retrieved from:
https://www.eca.europa.eu/Lists/ECADocuments/SR15_24/SR_VAT_FRAUD_EN.pdf

European Parliament. (2020*). Possible Solutions for Missing Trader Intra-Community Fraud*. Policy Department for Budgetary Affairs, Directorate-General for Internal Policies.

Fearing, A., Hopes, J., Jaras, T., Adeniji-Adele, A., Brooks, I., Kandemir, & Patel, K. (2015*). Implementing the destination principle to intra-EU B2B supplies of goods (Vol. 319). Final report* TAXUD/2013/DE.

Fedeli, S., & Forte, F. (2009). Models of cross-border VAT fraud. *Dipartimento Di Economia Pubblica, Working Paper n.123*.

Fiscalis. (2016*). The Concept of Tax Gaps: Report on VAT Gap Estimations*. The FISCALIS Tax Gap Project Group (FPG/041). Brussels, March 2020. Published by European Commission Director General of Taxation.

Fiscalis. (2018). *The Concept of Tax Gaps: Report III: MTIC Fraud Gap estimation methodologies*. The FISCALIS Tax Gap Project Group (FPG/041). Brussels, November 2018. Published by European Commission Director General of Taxation.

Fiscalis. (2022). MTIC VAT FRAUD. *Fiscalis Project Group on Estimations of Tax Gap On PIT/SSC, CIT, MTIC Fraud and VAT E-Commerce (FPG/008/001)*.

Frunza, M. (2016). *Cost of the MTIC VAT fraud for European Union members.* Available at SSRN 2758566.

Frunza, M., Guégan, D., & Lassoudière, A. (2010). *Statistical evidence of tax fraud on the carbon allowances market*. Available at SSRN 1664111.

Gebauer, A., & Parsche, R. (2003). Evasion of value-added taxes in Europe: IFO approach to estimating the evasion of value-added taxes on the basis of national accounts data (NAD). *CESifo DICE Report, 1*(2), 40-44.

Gebauer, A., Nam, C. W., & Parsche, R. (2005). VAT evasion and its consequence for macroeconomic clearing in the EU. *FinanzArchiv/Public Finance Analysis*, 462-487.

Gradeva, K. (2014). VAT fraud in intra-EU trade. *Goethe University, 33*.

Hamanaka, S. (2012). Whose trade statistics are correct? Multiple mirror comparison techniques: a test case of Cambodia. *Journal of Economic Policy Reform*, *15*, 33-56.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer.

HMRC. (2019). *Measuring Tax Gaps 2019 Edition*. London: HM Revenue & Customs.

HMRC. (2022). *HMRC Internal manual. VAT Fraud.* VATF23550.

House of Lords. (2007). *Stopping the Carousel: Missing Trader Fraud in the EU. Report with Evidence.* HL Paper 101.

Hudec, M., & Juriová, J. (2013). "Perspectives of data mining in improving data collection processes in official statistics." *Advances in Methodology and Statistics* 10.2: 65-81.

Jakobsson, O. (2013). *Missing trader fraud in the EU.* Faculty Of Law. Lund University.

Joński, K., & Gajewski, D. J. (2022). VAT Gap Estimation: Distinguishing Between Informality and Fraud. *EC Tax Review, 31*(3).

Jovanovic, Branimir. (2015*). Kalman Filter Estimation of the Unrecorded Economy in Macedonia.* 10.13140/RG.2.2.11649.61288.

Karaboytcheva, M. (2020). Addressing the VAT Gap in the EU. *EU Policies - Insights*.

Karanfil, Fatih and Ata Ozkaya (2007). Estimation of real GDP and unrecorded economy in Turkey based on environmental data. *Energy Policy, 35*, 4902–4908.

Kirchgässner, Gebhard (2016). On Estimating the Size of the Shadow Economy. *Economics Working Paper Series 1603*, University of St. Gallen, School of Economics and Political Science.

Kľúčik, M. (2012). "Estimates of foreign trade using genetic programming." *Proceedings of the 46 the scientific meeting of the Italian Statistical Society*.

Lindholm A. (2014). *A study about fraud detection and the implementation of SUSPECT - Supervised and UnSuPervised Erlang Classifier Tool.*

Loureiro, A., Torgo, L., & Soares, C. (2004). "Outlier detection using clustering methods: a data cleaning application." *Proceedings of KDNet Symposium on Knowledge-based Systems for the Public Sector.* Bonn, Germany.

Malan, J., Bosch Chen, I., Lamensch, M., & Pavesi, S. (2022). *Possible Solutions for Missing Trader Intra-Community Fraud*. Policy Department for Budgetary Affairs, Directorate-General for Internal Policies.

Mazurek, J., & Strzałka, D. (2022). On the Monte Carlo weights in multiple criteria decision analysis. *PLOS ONE, 17*(10), e0268950.

Nam, C. W., Parsche, R., & Schaden, B. (2001*). Measurement of value-added tax evasion in selected EU countries on the basis of national accounts data.* Available at SSRN 267677.

OECD (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*.

Olofsdottir, A. (2018*). Report: MTIC Fraud Investigation and LEA's cooperation improving.* Europol.

Polanec, S., Smith, P. A., & Bavdaž, M. (2022). "Determination of the Threshold in Cutoff Sampling Using Response Burden with an Application to Intrastat." *Journal of Official Statistics (JOS)* 38.4

Poniatowski, G., Śmietanka, A., & Bonch-Osmolovskiy, M. (2020). *Study and reports on the VAT gap in the EU-28 member states: 2020 final report.* CASE Research Paper, (503).

PricewaterhouseCoopers (2010). *Study on the feasibility of alternative methods for improving and simplifying the collection of VAT through the means of modern technologies and/or financial intermediaries. Final report*.

PricewaterhouseCoopers (2022*). VAT in the Digital Age proposals published by the European Commission. Tax Policy Alert*.

Reckon, L. L. P. (2009*). Study to quantify and analyse the VAT gap in the EU-25 Member State*s, Report for DG Taxation and Customs Union 20: 5–109.

Republic of Slovenia Statistical Office (2019). *Annual Quality Report for the Survey Intrastat – Trade in Goods Between EU Member States*.

Ruffles D., Tily G., Caplan D., & Tudor S. (2003). VAT missing trader intra-Community fraud: the effect on Balance of Payments statistics and UK National Accounts. *Economic Trends, No. 597*.

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond.* Cambridge, MA: MIT Press.

Sergiou, L. (2012). Value added tax (VAT) carousel fraud in the European Union. *The Journal of Accounting and Management, 2*(2).

Smith, S., & Keen, M. (2007). *VAT fraud and evasion: what do we know, and what can be done?* IMF Working Paper No. 07/31.

Sokanovic, L. (2017). Missing trader fraud as part of organised crime in EU. *Economic and Social Development: Book of Proceedings*, 160-168.

Stiller, W., & Heinemann, M. (2019). Do More Harm than Good? The Optional Reverse Charge Mechanism against VAT Fraud. *Preprint on Research Gate*.

Szabó B. (2019). Grouping of Misconduct Types in Case of VAT Fraud. *Magyar Rendészet, 2019/1*, 101—116.

Vanhoeyveld, J., Martens, D., & Peeters, B. (2020). Value-added tax fraud detection with scalable anomaly detection techniques. *Applied Soft Computing, 86*, 105895.

Vaškovič, M., Zídková, H., & Arltová, M. (2021). The Volume of MTIC Fraud between Poland and Czechia in Electronic Devices Trade: General Method of Carousel Fraud Estimation. *International Journal of Economic Policy in Emerging Economies, 1*(1).

Zídková, H., & Balíková, K. (2021). The definitive VAT system and its impact on tax collection. *European Journal of Government and Economics, 10*(2), 146-166.

Appendix A. Glossary of terms

Acquisition fraud is the simplest form of MTIC fraud, consisting of two actors: a conduit company and a missing trader, located in a different Member State. Under this scheme the conduit company supplies goods or services to the missing trader at a 0% VAT rate. The missing trader then supplies these goods to a customer in their Member State, collecting VAT. Rather than remitting this VAT to the relevant authorities, the missing trader then disappears, resulting in a revenue loss for their Member State equal to the VAT charged.

Broker companies are one of the actors forming part of MTIC fraud schemes, situated at the “end” of the fraudulent transaction chain. They acquire the supplies from either the missing trader or a buffer company in the chain and sell them to a business located in another Member State.

Buffer companies are normal traders added between the missing trader and broker company in order to delay detection. A given scheme can consist of more than one buffer company, which may or may not be aware of the fraud taking place.

Carousel fraud builds on the basic acquisition fraud model, but following acquisition from the conduit company, the missing trader supplies them domestically to a broker company (rather than the final consumer). This company then sells the goods back to the original conduit company at a 0% VAT rate and has its VAT repaid by the Member State. This process is then repeated (hence the name “carousel”), with each “turn” resulting in another loss of VAT revenue from the Member State of the missing trader.

A **conduit company** is a business supplying goods or services subject to fraud to the missing trader, located in a different Member State, which may or may not be aware of the fraud taking place.

Contra-trader fraud is a complex fraud scheme which consists of not one but two transaction chains – parallel to the fraudulent chain, a legitimate transaction chain is introduced, hindering detection. At its centre is a contra trader company, which participates in both chains. The missing trader performs an Intra-Community acquisition (0% VAT rate) from a conduit company, supplies the goods domestically to the contra trader and then disappears without remitting VAT. The contra trader then makes an Intra-Community supply to a company in a third Member State (the customer) and reclaims the VAT it paid upon acquisition from the missing trader. Simultaneously to this, the contra trader will make another Intra-Community acquisition directly from a different conduit company and sell them domestically to a broker company, which then makes an Intra-Community supply to a final customer in another Member State. Under this scheme, the contra trader is able to offset its input VAT (fraudulent chain) with the output VAT (legitimate chain), allowing it to feign legitimacy and minimise its liabilities.

Cost-Benefit Analysis is a tool used to evaluate the benefits against the potential costs (or risks) of an approach or decision and decide if the former outweigh the latter. This approach takes into account quantifiable factors, such as monetary costs, which allow direct comparison.

Under the **cross-invoicer fraud** scheme, the missing trader does not disappear immediately and instead employs fictitious invoices in order to avoid detection. These invoices either do not correspond to the actual movement of goods or are used to offset the liabilities of the missing trader. As in other MTIC fraud schemes, the missing trader acquires the goods from a company located in another Member State and supplies them to a domestic customer. However, this transaction chain is not reported and is replaced by a fictitious one, according to which the missing trader acquired the goods domestically (from a false or hijacked company) and then supplied them to a conduit company in another Member State – allowing them to request a VAT repayment from their Member State.

EC sales list returns are a reporting requirement of business making Intra-Community Supplies, and in some Member States also in the case of Intra-Community acquisitions. Following the introduction of EU VAT quick fixes in 2020, they are also needed in order for business to qualify for a zero VAT rate. The frequency at which they are submitted depends on the thresholds established in each country. EC sales lists typically include information such as the customer's name and EU VAT number, the country code, and the value of the reported transactions.

Intrastat is an EU trade data collection system, offering the most detailed statistics on Intra-Community trade. All EU VAT-registered businesses are required to file Intrastat declarations when moving goods between Member States, provided they are making Supply or Acquisition surpassing a certain threshold each year (which varies across Member States, years, and sectors). The information contained in these declarations also varies between Member States, but generally includes a description of goods, the Member State of dispatch/arrival, the CN8 code, and the quantity and value of goods.

A **missing trader** is a VAT registered person who acquires (or purports to acquire) goods or services without paying VAT, sells said goods or services with VAT, but ultimately does not remit the collected VAT to the appropriate national authority and instead disappears. In the EU context, the missing trader performs Intra-Community acquisition, exploiting the EU rule stating cross-border movement of supplies is VAT free.

Multi-Criteria Analysis is a flexible tool used for decision-making, which compares alternative options along a set of pre-determined criteria. Unlike Cost-Benefit Analysis, this approach allows both for quantitative and qualitative criteria, and assigns them subjective (opinion-based) weights.

The **reverse charge mechanism** moves the responsibility of reporting a VAT transaction from the seller to the buyer of a good or service. In practice this means that, rather than being charged VAT by the supplier, the acquirer accounts for it in his VAT return. This is a derogation of the general rule under which it is the supplier who is liable to pay VAT.

Within the **Standard Cost Model**, costs are calculated by multiplying the average cost of the required activity ("Price") by the total number of activities performed per year ("Quantity"). The price is a product of the average hourly labour cost (including prorated overheads) and the time required per action, while the quantity is the frequency of required actions multiplied by the number of entities concerned. Where appropriate, other types of costs, such as the cost of outsourcing or equipment, are taken into account.

Appendix B. Supplementary information to mapping MTIC fraud pathways

Confidential. Excluded from the published version of the report.

Appendix C. Principles behind the framework and its parameters

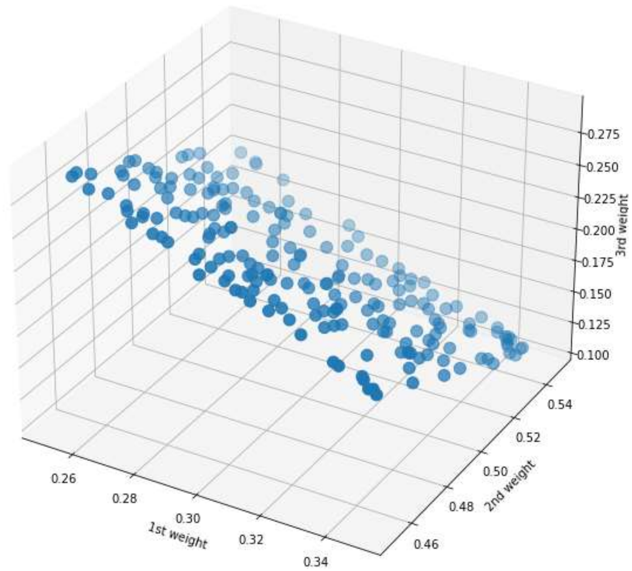
1) **Selection shall reflect the actionability of the study for the Commission and Member State administrations, and its informativeness for the broader public.** The objectives set by the studies are to: (1) help understand the size of this particular component of non-compliance; (2) monitor its evolution in time; (3) understand the nature and components of forgone revenue; and (4) provide insight into which strategies and measures are effective in reducing the scale (by looking at its evolution time, cross-country variation, and components). To reflect these objectives to the largest extent possible, the framework and its parameters were set in close cooperation with the European Commission and take into account the insights of the interviewed experts regarding potential goals of tax gap studies.

2) **Minimisation of the impact of subjective decisions by**:

   a) Accounting for uncertainty around framework parameters. The construction of the assessment methodology must involve a degree of subjectivity, as the MTIC gap study consists of multiple objectives, the relative importance of which could be perceived differently among Team members. As there is uncertainty around the values of weights, the Team proposes to run a Monte Carlo simulation,[117] in which values of these parameters are treated as a random variable. Following Mazurek and Strzałka (2022), we will consider sets of all feasible weights (a subspace of a *6*-dimensional space, where *6* is the number of sub-criteria), from which a large number of weights in the form of *6*-tuples is drawn randomly from a uniform probability distribution via the Monte Carlo method. The Team will then apply predefined dominance relations for the comparison and ranking of alternatives, and provide an analysis of the robustness of the rank of methodological approaches derived from a simulation of the aforementioned solutions. This simulation, based on a simpler, three-dimensional problem, is visualised in Figure 30, where coordinates in the scatter plot denote example weights in a three-criteria problem.

[117] In general, the term "Monte Carlo method" refers to a broad variety of algorithms that obtain numerical results via (many times) repeated random sampling from a given probability distribution (see Mazurek and Strzałka, 2022).

Figure 30: Visualisation of weights' combinations in a Monte Carlo simulation



Source: own elaboration.

b) **By quantifying the criteria to the highest possible degree.** The study team will quantify the criteria, including cases where this is only partially possible, using all available means. For example, when calculating the accuracy of the approaches, the part of the error that could be quantified (e.g., confidence intervals/standard errors in econometric-based methods) will be summarised using proper statistical techniques. Meanwhile, the part of the error that cannot be accurately quantified (e.g., human error) will be assessed by assigning its maximum and minimum values, and placing the value in broader intervals, based on experts' assessment.

3) **Completeness of the criteria.** Non-inclusion of some important criteria could lead to the assessment process straying from the objectives of the study. Such a framework could lead to an inaccurate ranking, for instance by giving a higher placing to approaches that would have been ranked poorly by the omitted criteria. To minimise the risk of incompleteness of the evaluation, we proposed a broad set of 11 sub-criteria (see Table 6).

4) **Minimising overlapping information**. If the criteria in the evaluation were largely overlapping, the risk of overcomplicating the framework and the identification of proper framework parameters would be elevated. Examples of inter-related criteria that were not included in tandem are the criterion of *complexity* of the study and the *expected cost*. Although these criteria are not fully overlapping, information gains from including both of them were expected to be marginal.

Focus on feasible solutions. The study team preselected the methods to assess only the approaches that are not clearly dominated and are feasible (now or going forward).[118]

---

[118] *Feasibility is interpreted in a broad sense covering also methodologies that could be, for instance, implemented by Member States rather than the Commission and its contractors.*

Appendix D. Analytical methods – supplementary information

Structural equation modelling (SEM) is a multivariate statistical analysis technique that is used to analyse structural relationships and estimate latent variables. This technique might be regarded as a combination of factor analysis and multiple regression analysis.

A specific example of the application of SEM is the Multiple Indicators, Multiple Causes Measurement (MIMIC) approach, which is one of the most popular methods for estimating the scale of the underground economy. The idea behind it is that the scale of the underground economy is a latent variable or index, which has causes and effects that are observable, but which cannot itself be directly measured. Thus, there are two kinds of observed variables in the model, "causal" variables, and "indicator" variables, which are connected by a single unobserved index. The values of the index over time are inferred from data on causes and indicators by estimating the statistical model and predicting the index. The fitted index is then interpreted as a time series estimate of the magnitude of the underground economy. Usually, the measure is hidden output or income as a percentage of recorded GDP, although some researchers are interested in the "tax gap" between actual and potential revenue when all taxable income is reported (Breusch, 2005).

There are three major reasons for using latent variable models. First, the approach is parsimonious because these models can summarise information contained in many response variables using only a few latent variables. Secondly, when properly specified, a latent variable model can minimise the biasing effects of measurement errors on estimating treatment effects. This property often makes this approach more accurate than a traditional version of the same analysis. Thirdly, latent variable models investigate effects between primary conceptual variables rather than between any particular set of ordinary response variables. Consequently, a latent variable model is often viewed as more theoretically appropriate than a simpler analysis with response variables only (Breusch, 2005, direct citation).

This use of structural equation modelling of the underground economy has its critics. According to Kirchgässner (2016) MIMIC is a confirmatory, rather than exploratory, statistical technique. It is not valid to conclude that a variable has been found as a statistically significant determinant of the shadow economy. As pointed out by Dybka et al. (2019), there is also a strand of research showing that the results of the MIMIC tend to be unstable. In addition, there is no economic theory to guide the specification, and the complexity of the estimation strategy is also often criticised. Moreover, the underground economy is not a latent or hypothetical quantity like intelligence; it is all too real, just difficult to measure (Breuch, 2005).

Kalman filtering uses a system's dynamic model, known control inputs to that system, and multiple sequential measurements to form an estimate of the system's varying quantities that is better than the estimate obtained by using a single measurement. The Kalman filter deals with the uncertainty of the noisy indicators described by observation equations and external random elements. It produces an estimate of the state of the system as an average of the system's predicted state and of the new measurement using a weighted average. These weights are calculated from the covariance, a measure of the estimated uncertainty of the prediction of the system's state. The result of the weighted average is a new state estimate that lies between the predicted and measured state and has a better estimated uncertainty than either alone. This process is repeated iteratively, with the new estimate and its covariance informing the prediction used in the subsequent iteration.

There are at least three studies that applied the Kalman filter technique to estimate the scale of the underground economy. Jovanovic (2015) used the filter to estimate the underground economy in Macedonia; Karanfil and Ozcaya (2007) – in Turkey; and Arango et al. (2006) – in Colombia.

As highlighted by Jovanovic, the Kalman filter has some conceptual advantages over the MIMIC approach. Unlike the latter, it does not model the unrecorded economy as a latent variable, but only as unobserved. Furthermore, it produces a direct estimate of the size of the unrecorded economy, while the MIMIC approach produces only an index of it, which then has to be transformed into cardinal values, assuming some value for the unrecorded economy for some period.

Appendix E. Responses to the questionnaire

Table 32: Responses to the questionnaire for tax administrations[119]

Member State	Belong to Fiscalis MTIC group	Questionnaire completed	MTIC fraud detection	Own estimates of the VAT revenue lost due to MTIC fraud	Provided (recent) estimates of VAT revenue lost to MTIC fraud[120]
Austria	No	Yes	Yes, annual	No	No
Belgium	No	Yes	Yes, annual	Yes, annual	No
Bulgaria	No	Yes	Yes, annual	Yes, but not annual	Yes
Croatia	Yes	Yes	No	Don't know	No
Cyprus	No	No			No
Czechia	No	Yes	No	No	No
Denmark	No	No			No
Estonia	No	Yes	Yes, annual	Yes, annual	Yes
Finland	Yes	No			No
France	No	Yes	No	No	No
Germany	No	Yes	Yes, annual	No	No
Greece	Yes	Yes	Don't know	Don't know	No
Hungary	Yes	Yes	Yes, annual	Yes, annual	No
Ireland	No	Yes	Yes, annual	No	No
Italy	Yes	Yes	Yes, annual	Yes, annual	No
Latvia	Yes	Yes	Yes, annual	No	No
Lithuania	Yes	Yes	Yes, annual	No	No
Luxembourg	No	No			No
Malta	Yes	Yes	No	No	No
Netherlands	No	Yes	Yes, annual	No	No

[119] Together with the European Commission, a decision shall be made on whether or not to include information on individual countries' MTIC calculation and detection efforts in the published Final Report.

[120] The answers in this column indicate whether or not the respondent provided estimates in the questionnaire itself, and may not reflect a Member State's ability to share these estimates with the Team (e.g., the respondent was not in a position to share them because they belong to a different authority).

Member State	Belong to Fiscalis MTIC group	Questionnaire completed	MTIC fraud detection	Own estimates of the VAT revenue lost due to MTIC fraud	Provided (recent) estimates of VAT revenue lost to MTIC fraud[120]
Poland	Yes	Yes	Yes, annual	Yes, but not annual	No
Portugal	No	Yes	No	No	No
Romania	Yes	No			No
Slovakia	Yes	Yes	No	No	No
Slovenia	No	Yes	Yes, annual	No	No
Spain	Yes	Yes	Yes, annual	No	No
Sweden	No	No			No

Source: Own elaboration.

Appendix F. Draft questionnaire for tax and statistical authorities

VAT compliance gap due to Missing trader intracommunity (MTIC):

Questionnaire for tax and statistical authorities

Fields marked with * are mandatory.

Introduction

This questionnaire forms part of the study *VAT compliance gap due to Missing trader intracommunity (MTIC),* commissioned by the European Commission, DG TAXUD.

The objective of the study in to select the methodological approach to estimating the forgone revenue due to MTIC fraud, and to implement it. The study is divided into two parts. During the first part, the study team researches previously used methodologies, review them according to the requirements of the European Commission and come up with a common methodology, which could be applied to all or the majority of Member States. Your knowledge and experience in MTIC gap calculations, together with your view on the accuracy of alternative methodologies would greatly benefit this part of the assignment and contribute to the shape of the final methodology.

All information shared will be treated as strictly confidential and will be processed only by the study team and the European Commission. The information will be used solely for the estimation of MTIC gaps and their components. The results of the analysis will be reported only in aggregate and none of the underlying figures will be made available to the public.

The deadline for the submission of responses is March 31, 2023.

If need be, the study team would be happy to organize a virtual meeting to explain and further discuss this questionnaire. If you would like to join such a meeting, or have any questions regarding this questionnaire, please do not hesitate to contact us (email: grzegorz.poniatowski@case-research.eu).

Thank you in advance for your support!

About You

A.1 First name (*)

A.2 Surname (*)

A.3 Email (*)

A.4 Name of your authority and organizational unit (*)

A.5 Your position in your national administration or authority (*)

A.6 EU member states (*)

- ☐ AT – Austria
- ☐ BE – Belgium
- ☐ BG – Bulgaria
- ☐ HR – Croatia
- ☐ CY – Cyprus
- ☐ CZ – Czechia
- ☐ DK – Denmark
- ☐ EE – Estonia
- ☐ FI – Finland
- ☐ FR – France
- ☐ DE – Germany
- ☐ EL – Greece
- ☐ HU – Hungary
- ☐ IE – Ireland
- ☐ IT – Italy
- ☐ LV – Latvia
- ☐ LT – Lithuania
- ☐ LU – Luxembourg
- ☐ MT – Malta
- ☐ NL – Netherlands
- ☐ PL – Poland
- ☐ PT – Portugal
- ☐ RO – Romania
- ☐ SK – Slovakia
- ☐ SI – Slovenia
- ☐ ES – Spain
- ☐ SE – Sweden

☐ I agree with the personal data protection provisions (*).

Data protection provisions.pdf

Availability of national estimates and related experiences

Q.1 Do you or other authorities in your Member State prepare (or have prepared in the past) own estimates of the VAT revenue lost due to **MTIC** fraud in your Member State? (*)

- ☐ No
- ☐ Yes – regularly, every year
- ☐ Yes - but not every year, and/or as one-off project(s)
- ☐ Yes - but it was discontinued
- ☐ Don't know

Q.2 Do you or other authorities in your Member State conduct (or have conducted in the past) **MTIC** fraud detection (risk assessment) of companies in your Member State? (*)

- ☐ No
- ☐ Yes - regularly every year
- ☐ Yes - but not every year, and/or as one-off project(s)
- ☐ Yes - but it was discontinued

Q.3 Please indicate which institution (or institutions, if more than one) is, or was, responsible for the measurement or detection of MTIC fraud in your Member State

Q.4 In which year did your Member State begin measuring or detecting MTIC fraud?

Q.5 If the size of MTIC fraud was measured, please list the years for which those estimates were produced and include relevant estimates (in nominal terms, as percent of VAT revenue or liability).

Q.6 What methodology or methodologies were used for the measurement/detection of MTIC fraud? Multiple answers are allowed.

- ☐ **Top-down estimation methods**:
  - ☐ Based on **trade value or balance**
  - ☐ Based on **trade mirror statistics**
  - ☐ Based on **data on VAT refunds and deductions**
  - ☐ Employing **econometric modelling**
  - ☐ Employing **machine learning**
  - ☐ Using **basic algebraic operations** on time series and panel data
- ☐ **Bottom-up estimation methods:**
  - ☐ Based on **random audits**

- ☐ Risk-analysis system – **traditional risk scoring**
- ☐ Risk-analysis system – **employing machine learning techniques**
- ☐ Methods based on matching data from **EC sales lists with VAT returns**

☐ Calculating **discrepancies in the fiscal data**

☐ **Other** methodologies and/or data sources (please describe below):

Q.7 Which of the following dimensions of MTIC measurement could be distinguished within the methodology applied in your Member State? Multiple answers are allowed.

- ☐ Type of fraud (e.g., carousel fraud, acquisition fraud)
- ☐ Trade partner (Member State which was at the other end of the fraudulent transaction)
- ☐ Type of taxpayer (e.g., individuals, companies, company types)
- ☐ Industries or types of goods or services which were the object of fraud
- ☐ Other (list all):

- ☐ The estimation was not split into any dimensions, only total value was estimated

Q.8 How much time and how many team members were required the first time the measurement/detection of MTIC fraud was undertaken?

Please include the time needed for the modelling and gathering of additional information, if applicable. The time can be reflected in FTEs (full-time-equivalents) or, if preferred, described in another way (e.g., how many people were involved in the task, what was the length of the project, what was the average engagement of team members)

Number of team members involved	Duration of the estimation	Average level of engagement

<u>or</u>

FTE

Q.9 How much time and how many team members were required to update the results in subsequent years?

The time can be reflected in FTEs (full-time-equivalents) or, if preferred, described in another way (e.g., how many people were involved in the task, what was the length of the project, what was the average engagement of team members)

Number of team members involved	Duration of the estimation	Average level of engagement

or

FTE

Q.10 What datasets are used as **primary sources of information** for estimation/detection of MTIC fraud in your Member State? Are these datasets prepared specifically for this purpose?

Q.11 Please express your view on the **accuracy of various alternative approaches** to measurement/detection of MTIC fraud - be it at the national or EU level.

	Unreliable/ inaccurate	Reliable, but prone to inaccuracies	Reliable/ accurate	Do not know
Top-down estimation methods:				
Based on trade volume and balance	☐	☐	☐	☐
Based on trade mirror statistics	☐	☐	☐	☐
Based on data on VAT refunds and deductions	☐	☐	☐	☐
Employing econometric modelling	☐	☐	☐	☐
Employing machine learning	☐	☐	☐	☐
Using basic algebraic operations with time series and panel data	☐	☐	☐	☐
Bottom-up estimation methods:				
Based on random audits	☐	☐	☐	☐
Risk-analysis system – traditional risk scoring	☐	☐	☐	☐
Risk-analysis system – employing machine learning techniques	☐	☐	☐	☐
Methods based on matching data from EC sales lists with VAT returns	☐	☐	☐	☐

	Unreliable/ inaccurate	Reliable, but prone to inaccuracies	Reliable/ accurate	Do not know
Calculating discrepancies in the fiscal data	☐	☐	☐	☐
Other methodologies and/or data sources (please describe below)	☐	☐	☐	☐

Please describe the other method:

Q.12 Please indicate which sources of information are available in your Member State (for the period 2017-2021).

	Readily available for the whole period	Available but require additional work	Partially available (e.g., only some years)	Not available	Do not know
Consolidated **random audit results**	☐	☐	☐	☐	☐
Consolidated **risk-based audit results**	☐	☐	☐	☐	☐
Value of estimated size of **MTIC fraud**	☐	☐	☐	☐	☐
List of goods and services (or industries) identified as most commonly used for MTIC fraud	☐	☐	☐	☐	☐
List of **preventive measures introduced** in an attempt to counter MTIC fraud	☐	☐	☐	☐	☐
Aggregated data on VAT refunds	☐	☐	☐	☐	☐
Aggregated data on VAT deductions					
Internal documentation on the implementation of MTIC fraud measurement/detection solutions	☐	☐	☐	☐	☐
Reports summarizing the implementation of MTIC fraud measurement/detection solutions	☐	☐	☐	☐	☐
Other relevant data used for MTIC fraud measurement or detection (describe in lines below):					
______________________	☐	☐	☐	☐	☐
______________________	☐	☐	☐	☐	☐
______________________	☐	☐	☐	☐	☐

Q.13 Please indicate which sources of information can be shared by your Member State (for period 2017-2021).

	Could be shared in entirety	Can be shared in parts	Cannot be shared
Consolidated **random audit results**	☐	☐	☐
Consolidated **risk-based audit results**	☐	☐	☐
Value of estimated size of **MTIC fraud**	☐	☐	☐
List of goods and services (or industries) identified as most commonly used for MTIC fraud	☐	☐	☐
List of **preventive measures introduced** in an attempt to counter MTIC fraud	☐	☐	☐
Aggregated data on VAT refunds	☐	☐	☐
Aggregated data on VAT deductions			
Internal documentation on the implementation of MTIC fraud measurement/detection solutions	☐	☐	☐
Reports summarizing the implementation of MTIC fraud measurement/detection solutions	☐	☐	☐
Other relevant data used for MTIC fraud measurement or detection (describe in lines below):			
______________________	☐	☐	☐
______________________	☐	☐	☐
______________________	☐	☐	☐

Q.14 Do you think it would be technically possible to **match data** on Intra-Community Supply/Acquisition from **EC Sales Lists** (from the VIES system) with corresponding deliveries in VAT returns?

☐ Yes

☐ No

Q.15 If yes, please explain what are the potential problems with **matching** data on Intra-Community Supply/Acquisition from **EC Sales Lists** (from the VIES system) with corresponding deliveries in VAT returns?

Q.16 Could following information be potentially shared with the European Commission and the contractor?

	Yes	No	If no, what is the reason
Anonymized individual data from VIES and Intra-Community acquisitions from tax returns	☐	☐	
Value of unmatched Intra-Community Acquisitions from VIES and tax returns	☐	☐	

Aggregate Intra-Community Acquisitions from VIES and tax returns	☐	☐	

Q.17 Do you have any comments regarding the project, this questionnaire, or specific questions within it? Please share them below.

Q.18 If you can share any documents, reports or any other relevant materials on MTIC fraud measurement, please upload the files here. They would be of great use to our project.

Appendix G. Selected revealed cases of MTIC

Table 33: Selected revealed cases of MTIC

Member State	Year of press release	Group of products and services	Approximated value of forgone revenue (EUR mln)	Source
Poland	2023	electronic devices	350	https://www.gov.pl/web/kas/karuzela-podatkowa-w-obrocie-elektronika-10-osob-zatrzymanych
Italy	2023	electronic devices	28	https://www.eppo.europa.eu/en/news/italy-eppo-arrests-suspected-ringleader-eu28-million-vat-fraud-scheme
Italy	2023	electronic devices	40	https://www.eppo.europa.eu/en/news/eppo-uncovers-eu40-million-vat-fraud-six-arrests-and-seizures-sting-against-organised-crime
Italy	2023	liquid fuels	92	https://www.eppo.europa.eu/en/news/italy-eppo-seize-eu149-million-investigation-vat-fraud-fuels
Spain	2023	electronic devices	25	https://www.eppo.europa.eu/en/news/eppo-busts-eu25-million-vat-fraud-spread-across-eight-countries-17-arrests-including-alleged
Austria, Bulgaria, Czechia, Estonia, Germany, Hungary, Lithuania, the Netherlands, Slovakia and Slovenia	2023	electronic devices	32	https://www.eppo.europa.eu/en/news/eppo-investigation-cross-border-vat-fraud-estimated-damages-eu32-million-61-searches-10
Belgium, Cyprus, France, Germany, Greece, Hungary, Italy, Lithuania, Luxembourg, the Netherlands, Portugal, Romania, Slovakia and Spain	2022	electronic devices	2,200	https://www.eppo.europa.eu/en/news/operation-admiral-eppo-uncovers-organised-crime-groups-responsible-vat-fraud-estimated-eu22
Poland	2022	cooking oil	38.9	https://www.gov.pl/web/kas/karuzela-podatkowa-w-obrocie-olejem-rzepakowym-6-osob-zatrzymanych
Cyprus	2022	computers	-	https://knews.kathimerini.com.cy/en/news/cyprus-helps-uncover-vat-carousel-fraud
Germany	2022	platinum coins	-	https://www.eppo.europa.eu/sites/default/files/2023-02/EPPO_2022_Annual_Report_EN_WEB.pdf
Hungary, Germany, Italy	2021	new and used cars	38	https://www.europol.europa.eu/media-press/newsroom/news/five-suspects-responsible-for-eur-38-million-vat-fraud-scheme-arrested

Member State	Year of press release	Group of products and services	Approximated value of forgone revenue (EUR mln)	Source
Hungary	2021	mobile phones	29.8	https://www.europol.europa.eu/media-press/newsroom/news/%E2%82%AC142-million-seized-cross-border-vat-fraudsters-in-hungary
Hungary, Croatia	2021	services (not specified)	8	https://www.europol.europa.eu/media-press/newsroom/news/hungarian-authorities-break-%E2%82%AC8-million-vat-fraud-scheme
Netherlands	2021	memory cards for electronic devices	9	https://www.europol.europa.eu/media-press/newsroom/news/vat-fraud-clampdown-international-scam-memory-cards-uncovered-in-netherlands
Germany, Poland	2021	cooking oil	17.8	https://www.europol.europa.eu/media-press/newsroom/news/police-dismantle-criminal-network-linked-to-international-vat-fraud-trading-vegetable-oil
Poland	2021	cooking oil	37.8	https://www.gov.pl/web/kas/karuzela-podatkowa-przy-obrocie-olejem-rzepakowym-17-osob-z-zarzutami
Poland	2021	electronic devices	6.25	https://www.gov.pl/web/prokuratura-krajowa/prokuratura-regionalna-we-wroclawiu-areszty-dla-czlonkow-zorganizowanej-grupy-przestepczej-wyludzajacej-vat-w-ramach-karuzeli-podatkowej
Italy	2021	new and used cars	6.3	https://www.eurojust.europa.eu/news/eurojust-assists-italy-blocking-vat-fraud-scheme-car-imports-germany
Spain	2021	-	26	https://www.eurojust.europa.eu/news/eurojust-supports-spanish-action-against-massive-vat-fraud
Germany, Italy and Bulgaria	2021	luxury cars	13	https://www.eppo.europa.eu/en/news/international-strike-against-organised-crime-group-10-arrests-and-seizures-worth-least-eu13
Hungary, Croatia	2020	sugar and cooking oil	10	https://www.europol.europa.eu/media-press/newsroom/news/vat-scammers-arrested-in-hungary-after-evading-close-to-%E2%82%AC10-million-in-tax
Hungary, Slovenia	2020	soybeans	1.4	https://www.europol.europa.eu/media-press/newsroom/news/3-arrested-in-hungary-in-%E2%82%AC14-million-vat-fraud-investigation
Portugal, Germany,	2020	new and used cars	5	https://www.europol.europa.eu/media-press/newsroom/news/network-defrauding-least-%E2%82%AC5-million-dismantled-in-portugal
Spain, Portugal	2020	food and alcoholic beverages	-	https://www.europol.europa.eu/media-press/newsroom/news/cross-border-vat-fraudsters-busted-in-portugal-6-arrests-and-32-indictments
Poland	2020	construction, advertising and transport services	35.6	https://www.gov.pl/web/prokuratura-krajowa/kolejne-uderzenie-zachodniopomorskiego-pionu-pz-pk-w-zorganizowana-grupe-przestepcza-prowadzaca-tzw-karuzele-podatkowa
Poland	2020	FMCG and stretch film	3.3	https://www.gov.pl/web/prokuratura-krajowa/zatrzymanie-czlonkow-zorganizowanej-grupy-przestepczej-zajmujacej-sie-karuzelami-podatkowymi
Poland	2020	gold	2.2	https://www.gov.pl/web/prokuratura-krajowa/odpowiedza-przed-sadem-za-wyludzenia-podatku-vat-w-zwiazku-z-pozorowanym-obrotem-zlotem
Poland	2020	liquid fuels	70	https://www.gov.pl/web/kas/karuzela-vat-zlikwidowana-przez-kas-i-cbsp

Member State	Year of press release	Group of products and services	Approximated value of forgone revenue (EUR mln)	Source
Bulgaria, the Czech Republic, France, Germany, Spain, Latvia, Lithuania and Romania	2020	luxury cars	-	https://www.eurojust.europa.eu/news/eurojust-supports-major-crackdown-vat-fraud-car-sales
Hungary	2020	perfumes	8.5	https://www.eurojust.europa.eu/news/large-scale-vat-fraud-fictitious-export-perfumes-halted-hungary
Hungary, Austria	2019	mobile phones	12	https://www.europol.europa.eu/media-press/newsroom/news/carousel-of-vat-abuse-dozens-arrested-in-connection-multi-million-tax-evasion-schemes
Hungary	2019	-	70	https://www.europol.europa.eu/media-press/newsroom/news/over-%E2%82%AC70-million-seized-in-hungary-in-operation-backbone
Poland	2019	new and used cars	1.6	https://www.gov.pl/web/kas/karuzela-podatkowa-rozbita-przez-zachodniopomorska-kas
Czech Republic	2019	steel	12	https://www.novinky.cz/clanek/krimi-kraceni-dani-za-300-milionu-v-hutnim-byznysu-na-zlinsku-policie-navrhla-obzalobu-40303508
Italy, Slovenia	2019	marketing of technological and computer products	500	https://www.eurojust.europa.eu/news/eurojust-helps-italy-and-slovenia-unravel-massive-vat-fraud
Italy, Germany, Spain, Austria and Romania	2019	beer and non-alcoholic beverages	2	https://www.eurojust.europa.eu/news/vat-carousel-fraudsters-arrested-support-eurojust
UK	2019	renewables energy certificates of origin	-	https://www.theguardian.com/politics/2019/jun/19/hmrc-targets-fraudsters-taking-billions-in-renewable-energy-vat-certificates
Belgium, Bulgaria, Germany, Hungary, Italy, Portugal, Romania and Spain	2018	electronic devices	140	https://www.europol.europa.eu/media-press/newsroom/news/eu-wide-vat-fraud-organised-crime-group-busted
Poland	2018	electronic devices	15.6	https://www.gov.pl/web/prokuratura-krajowa/zatrzymania-zorganizowanej-grupy-przestepczej-zajmujacej-sie-popelnieniem-tzw-przestepstw-karuzelowych
Czech Republic	2018	steel	0.9	https://ct24.ceskatelevize.cz/regiony/2412220-z-prodeje-oceli-neodvedl-22-milionu-na-dph-soud-poslal-karuseloveho-podvodnika-na-65
Romania, Germany	2018	used cars	20	https://www.eurojust.europa.eu/news/international-action-against-large-scale-vat-fraud
Belgium, Bulgaria, Germany, Spain, Italy, Hungary, Portugal and Romania	2018	electronic devices	17	https://www.eurojust.europa.eu/news/international-vat-fraud-ocg-dismantled
Romania, Italy	2018	electronic devices	30	https://anti-fraud.ec.europa.eu/media-corner/news/finance-police-latvia-stops-group-involved-vat-fraud-and-money-laundering-2017-09-29_en
UK	2018	scrap metal	0.3	https://www.letsrecycle.com/news/carousel-fraud-concerns/
Poland	2017	liquid fuels	-	https://www.gov.pl/web/prokuratura-krajowa/areszty-w-sledztwie-dotyczacym-nieprawidlowosci-w-obrocie-paliwami-cieklymi

Member State	Year of press release	Group of products and services	Approximated value of forgone revenue (EUR mln)	Source
Poland	2017	fireproof stain and solar panels	1	https://www.gov.pl/web/prokuratura-krajowa/wyludzili-ponad-5-milionow-zlotych-vat-na-fikcyjnym-obrocie-bejca-ognioodporna-i-panelami-slonecznymi
Lithuania	2017	electronic devices	-	https://anti-fraud.ec.europa.eu/system/files/2021-07/20012017_srs_fpd_stops_2_vat_fraud_schemes_en.pdf
Germany	2015	electronic devices	57	https://www.europol.europa.eu/media-press/newsroom/news/eight-member-states-take-action-against-international-vat-fraud
Poland	2015	-	320	https://www.europol.europa.eu/media-press/newsroom/news/strong-collaboration-targets-vat-fraud
Poland, Germany	2015	oil and luxury cars	20	https://www.europol.europa.eu/media-press/newsroom/news/asset-recovery-successful-cooperation-between-europol-and-poland
Italy	2015	cooking oil	-	https://www.europol.europa.eu/media-press/newsroom/news/millions-of-euros-recovered-in-operation-against-excise-fraud
Italy	2014	alcoholic beverages	-	https://www.europol.europa.eu/media-press/newsroom/news/europol-and-eurojust-support-successful-action-against-alcohol-carousel-fraud
Hungary, Czechia, Slovakia	2013	scrap metal	2.7	https://www.europol.europa.eu/media-press/newsroom/news/four-million-euros-seized-14-suspected-vat-fraudsters-arrested
Most EU MS	2010	Carbon Emissions Trading System	5,000	https://www.europol.europa.eu/media-press/newsroom/news/carbon-credit-fraud-causes-more-5-billion-euros-damage-for-european-taxpayer

Source: Own elaboration.

Appendix H. Matrix with comments and responses

Inception Report

COMMENT	RESPONSE
Michael Udell District Economics	
Tables 2 and 3 of the draft identify 3 dimensions upon which MTIC VAT fraud methodologies are to be ranked. The first, and most restrictive, is comparability across EU member countries. Table 8 of the VAT Gap Report 2022 provides a starting point for this criterion because 27 of the 30 EU member countries and 24 of the 26 Schengen area countries already have VAT gap estimates for the policy, rate, exemption, and actionable gaps. Is it the case that the common basis for these estimates at the country level (see for example the estimates for Belgium on table 9) is a top-down consumption/production side VAT gap analysis? If so, can these estimates form the basis for cross-country comparisons for MTIC intracommunity VAT fraud? I believe the answer is yes. These estimates provide “control totals” for VAT non-compliance, of which MTIC VAT fraud is a component.	Indeed, the estimates could provide “control totals”. Yet, calculating the components of the gap is the main challenge that the study team will be faced with.
As a first step consider the tax policy characteristics shown on table 63 of the VAT Gap Report 2022. IT investment, verification interventions, electronic payments, and reporting obligations are good tax system measures but surprisingly provided little in the way of explanatory power. I encourage you to not give up on the concept of tax system parameters as explanatory variables for two reasons. First, these are policy variables that a government can adjust, and second, policy variable differentials across borders should result in differential tax revenue outcomes. A good place to start would be to use the tax rates and exemption levels of each country’s VAT system. Rate and exemption differentials across borders create a push-pull that attracts or repels buyers. One of the benefits of objective system measures is that the relevant parameters are easily available as tax system rates and exemptions and do not require separate analysis to create. These parameters tend to change over time as each country adjusts its tax parameters.[121] In this institutional parameter approach a null hypothesis might be “does the MTIC VAT fraud occur uniformly within the community, or are some borders more susceptible than others to this behavior?”	Once the estimates of the MTIC gap are available, it would be very informative to verify if MTIC VAT fraud occurs uniformly within the Community, or if some borders are more susceptible to this behaviour.

[121] *In the US, I live in the state of Maryland, which is generally regarded as a high tax state. We have a 6% sales tax. Our neighbour state, Delaware, has a 0% sales tax. Delaware attracts large crowds of weekend shoppers to buy items without sales tax.*

The challenge is to characterize the deviations from a notional uniform MTIC VAT fraud across each border. An example may help. Table 1 below shows each main component of VAT Total Tax Liability in the first column. Consider the 27 EU member countries listed in the VAT Gap Report 2022. For any two countries "I" and "j" there are VAT gap estimates for the components of VTTL. Would it be possible to calculate tax policy parameters around the gap estimates based on VAT rate differentials and exemption differentials? In its simplest form a rate differential would be a general VAT rate in country "I" minus the same in country "j" and an exemption differential might be the same in euros. The differentials could have signs indicating directionality. That is, if the general rate that households face in Finland is 24% and in Estonia 20%, then the rate differential could be -4% if we are trying to explain MTIC VAT fraud from Estonia to Finland, or +4% if we are trying to explain MTIC from Finland to Estonia. Deciding which VAT rates are applicable – the general rate or special rates – will require some judgement. One approach might be to consider the goods and services involved in the Finland-Estonia cross border trade using the World Bank trade tables to determine whether the items most traded would be subject to the general rates or the special rates.	As noted by the Reviewer, policy parameters are surely one of the main groups of factors explaining the scale of the MTIC gap (and its spatial dimension). Unfortunately, many of these 'parameters' cannot be quantified. One example of this is the effectiveness of tax administrations in tackling fraud. As a result, such a spatial model would likely suffer from omitted variable bias. Moreover, the feasibility of the proposed approach is likely limited, as we know only the control totals (the entire VAT compliance gap) and do not know how much of that should be attributed to MTIC fraud.
The basic idea (one we are developing extensively using gravity models) is that rate (and exemption) differentials across borders have a push-pull effect on economic flows. Gravity models of trade have been applied ever since Jan Tinburgen first applied it to trade in 1962.[122] In our research, rather than using measures of economic mass such as GDP or population, we use objective tax administration measures relevant to the topic that can describe attraction/resistance across a boundary between two countries. Because the topic is the VAT tax gap and MTIC intracommunity fraud, overlaying the VAT tax rate and exemption differentials across each of the borders in the community could provide a basic "topography" of attraction. In our research on gravity models in the cross-border income tax evasion context, we find evidence that this cross-border policy topography has predictive power with respect to certain tax flows without using any measures of economic mass.	The VAT rate is also likely one of the factors behind the scale of MTIC fraud, but probably not a major one. Our analysis of MTIC fraud pathways has shown that fraudsters choose specific Member State due to the inefficiency of their administrations (which is very difficult to determine and quantify) and the specificities of markets for selected goods in selected MS. The rate could probably explain only a small fraction of the MTIC variability, which could hopefully be verified once the estimates are available.

[122] See Jan Tinburgen, "An Analysis of World Trade Flows", in Shaping the World Economy, Twentieth Century Fund, New York, 1962.

Gathering this discussion, consider a model where the dependent variable is the household component of the VAT tax gap for a country and the independent variables are the cross-border tax differentials for each country that shares a border, the trade volumes across the same borders (or if available, the trade volumes of household goods across the same borders). We might expect such a model to have relatively low to modest fit but were these parameters to explain 10% of the variation in household VAT tax gap over the EU, then that would be a VAT tax gap estimate attributable to cross-border trade using the same top-down estimation approach used for the VAT Gap Report 2022. The usefulness of this approach is that it provides a consistent measure of cross-border attributable VAT tax gap. It may not specifically identify MTIC VAT fraud, which would be a component of cross-border VAT fraud.	Unfortunately, the household component of the VAT compliance gap cannot be quantified as there are no revenue figures that could be inter-related with this liability component (household consumption). For this reason, the consumption-based approach does not allow one to break the VAT compliance gap into components.
The second criterion to use for judging methods of estimating the MTIC VAT tax gap concerns the granularity of the analysis. Can an analytic method provide results that can be “linked” in some meaningful way back to an MTIC VAT gap estimate that satisfies the comparability criteria? This is important because there will be many different analyses of MTIC VAT gap issues depending upon country and tax administrator views about what types of MTIC schemes are challenging their tax systems. In my opinion, it is good for tax compliance to encourage each tax administration to focus on their understanding of the most pressing version of MTIC VAT fraud. Where the dimension of granularity comes in is” are there approaches and best practices that each tax administrator can ascribe to that allow for statistical linkage to country-level MTCI VAT fraud estimates that satisfy the comparability criteria?”	The importance of granularity was well-noted, and the assigned weight was elevated.
There is at least one large challenge to this approach towards allowing highly granular audit campaigns to serve as inputs to estimation of MTIC VAT fraud. Member countries vary greatly in size. If audit campaigns by small countries are unique across member countries, even if the campaign results could be statistically linked back to that country’s estimate of MTIC VAT fraud, it may not be informative for the community. This does not mean that tax administrators should not develop campaigns to address MTIC VAT fraud in the “flavor” most concerning to their country. It does mean that there may be a large country bias to allowing the results of one-off campaigns to be used in estimates of MTIC VAT fraud.	The issue of incomparability of audit results and efficiency of audits across Member State was well-noted and stated in the Interim Report.

Prof. Silvia Fedeli Sapienza - Universita' di Roma	
[…] I believe that the first issues to be addressed are related to the different technicalities used by fraudster in the different sector of economic activities, how much these technicalities are influenced by the commodity characteristics of the traded goods, and how different (i.e., non-harmonized) legislations of member states enable the perpetration of fraud. All these issues are quite important, not only in terms in terms of antifraud policies, but mainly to the end of both the evaluation of the existing methodologies and the policy requirement of improving these same methodologies. Right now the narrative of this chapter is very basic, and I would suggest -beginning with this chapter, but also to be deepened subsequently with the questionnaire- to **write a comprehensive taxonomy of MTIC fraud** that signals to the reader where the flaws lie and their likely reasons, for example, the variegate national legislations, or the commodity nature of the traded goods and services suitable for MTIC, or weaknesses of the overall system, or a combination of all these causes. In my view this should be the starting point of any empirical evaluation.	We fully agree that the technicalities of the fraud must be carefully scrutinized and reported. This report contains an extended background section, i.e., a full mapping of MTIC fraud pathways, based on the literature and interviews with practitioners from eight Member States. As advised, the literature review was extended by including papers that deal with the technicalities of the fraud (to the extent they are useful for selecting the methodology for estimating the fraud). Yet, the effectiveness of measure to tackle fraud is beyond the scope of our work.
I would suggest a survey of the literature not only on the existing methods used for MTIC gap estimation, but also on papers and documents based on the expertise of VAT practitioners explaining the technicalities of MTIC fraud. I noticed also that some relevant academic paper are missed, possibly because not empirically relevant. I strongly suggest the study team to be aware of the qualitative aspect of all the known specific MTIC fraud cases, better if related to specific economic activities. In principle, not knowing the object of the proposed estimate could invalidate the entire result of the analysis. All the more so in this case because of the nature of the phenomenon under analysis that per se can neither be directly observed nor fully disclosed by standard statistics.	See above.

This point (page 10 of the report) should be clarified. I am not sure of the content. It seems that there is an overlapping with the Fiscalis Project Group that seems very important/relevant for the study team and that, at the moment, has been neglected. I might have misunderstood.	The discussion on the effects of the work of the Fiscalis Project Group will be expanded after the meeting of the group on April 21.
As for the utility of the questionnaire, I can see one important issue to be disclosed in depth. It has to do with perceived accuracy of pre-selected method types […] What I rather suggest the questionnaire achieves is what has been excluded from MTIC fraud in the proposed alternative methodologies. I do not see the importance for the analysis of eliciting the vague opinion of the person responsible for the response in the suggested term. (i.e., Unreliable/ inaccurate, Reliable, but prone to inaccuracies, Reliable/accurate, Do not know). Note here that the alleged respondents are fairly well informed about the extent of the phenomenon in their country. Therefore, the respondents should be somewhat compelled to say what, in their opinion, was not captured by various alternative approaches to MTIC fraud measurement/detection and why.	As the authorities are very important stakeholders with hands-on experience, we need to make sure that their knowledge and perceptions are well accounted for. One of the main objectives is the actionability of the analysis for MS' administration, which cannot be achieved with tools that are not trusted.
I see no reason why follow-up interviews should be limited to 2-3 MSs only.	First, the study team has limited resources and timeframe, so the proposed scope of work must be in line with the proposal and the ToR. Second, the questionnaire was comprehensive, and the in-depth interviews shall only be needed with the administrations that have extensive experience with MTIC gap calculations.
Task 3 of this chapter broadly aims to develop a common methodology. This task is not very clear to me. The purpose (task) of mapping these "archetypal paths of MTIC fraud" is a rather obscure concept, since any eventual mapping should emerge from previous tasks 1 and 2. Moreover, the Study team says it wants to draw on the expertise of VAT practitioners to understand MTIC fraud pathways. As already mentioned, in my view, this should be done at the outset, to have a taxonomy of MTIC fraud clear in mind: I mean all the work determined by Box 1 should flow into Chapter I as a part of the background chapter of the entire report. Hence, task 3 should begin with the identification of the data to be used or suitable for estimation. Table 1 ("not exhaustive on the relevant sources of data on MTIC fraud, along with their availability")	Task 1 and 2 are defined by the ToR as *mapping of the methodologies and methods used by EU Member States, third countries, and international organisations, collection of information from existing documents*. Mapping of the fraud will build on these findings but is not a part of these tasks, thus it must be put under the umbrella of Task 3. The **structure of the Final Report will not build on the structure of the Inception Report**, which had to be submitted before the practitioners were interviewed. The

is a good start. Unfortunately, it seems to me that the most important data for the study are those not publicly available in the possession of national tax agencies -notice that this motivated my previous comment on the questionnaire-. The analysis should then follow with the evaluation framework to be decided upon and proposes some evaluation criteria.	Final Report will therefore include the mapping at the outset, rather than in the section discussing the next steps.
I am very concerned about the arbitrary weights given to both the criteria (and sub-criteria) and the pre-selection requirements (column 3) in Table 3. I am even more concerned given the proposed methodology suggested for comparison. […] According to me, for this very reason, although theoretically interesting, the joint CBA with the MAC may not be suitable for choosing a quantitative methodology. The latter, contrary to the study group's assertion, should not, per se, be prone to multiple objectives of the subject involved/interested. Rather, the quantitative methodology should be as accurate as possible. In this respect, I find that establishing subjective parameters for the choice such as (1) weighting and (2) scaling parameters is a way to introduce ex ante bias that might invalidate the analysis.	Could you please share some ideas on an evaluation framework that would not be based on pre-determined weights? In this Interim Report we propose a simulation that will test the sensitivity of the methodology selection with respect to the weights that were assumed. In our view, **accuracy cannot be the only objective of the study**. Otherwise, we might pick a method that covers only a single year, single country, does not allow for any breakdown, cannot be continued in the future, etc.
The study team does not make explicit how the Cost-Benefit Analysis (CBA) is to be carried out. This is rather disappointing given that, technically, Multi-Criteria Analysis should be linked to CBA. On the basis of what is said now in the report, the suggested combination of CBA and MCA, or the MCA alone, would be misleading since, as acknowledged by the study team, Multi-Criteria Analysis is an opinion based method.	This is explained in detail in Table 3. As explained in the introduction, we included this discussion earlier than envisaged by the project agenda.
However, in this regard, I also note that no mention is made of the important task of estimating MTIC fraud by sector of economic activity of the MSs. In terms of designing policies against VAT fraud on the basis of MTIC fraud evaluation, however, this topic is quite important and the Study Group ought to plan to seriously consider it in the study.	This is one of the reasons for which accuracy cannot be a single criterion. The assessment of the possibilities of breakdown will be an important subject of the next step of work.

Draft Final Report

COMMENT	RESPONSE
Prof. Silvia Fedeli Sapienza - Universita' di Roma	
[General] […] the Mapping of MTIC fraud pathways […] is based on an excessively specific and marginal case, deviating from the standard EU VAT (of consumption type) system based on the principle of destination. The provided basic example of MTIC carousel fraud, which serves as the foundation for all cases examined, offers a limited and peculiar perspective on the matter. It exclusively pertains to a very specific case involving the reverse charge mechanism, which is neither the norm nor a long-term practice within the EU. […] Considering that all cases examined in the present final report mistakenly assume a widespread utilization of the reverse charge mechanism (which is not the norm in the EU), the report itself is founded upon a significant underlying error that leads to a substantial underestimation of the phenomenon. Thus, I strongly recommend conducting a comprehensive review of all comments pertaining to this phenomenon across various chapters, based on this misrepresentation.	The reverse charge mechanism on intra-Community transactions was introduced with the launch of the single market in 1993 and shifts the responsibility for the recording of a VAT transaction from the seller to the buyer of a good/service, thus eliminating or reducing the obligation for sellers to VAT register in the country where the supply is made. If the supplier incurs any local VAT on costs related to the service or goods supplied under the reverse charge, they may recover them through an EU VAT reclaim. It is our belief that the fact that intra-community movement of goods is effectively VAT-free (and that buyers are responsible for recording the VAT both as input and output under the reverse charge mechanism) is the central to the existence of MTIC fraud specifically. Perhaps, you had in mind the *domestic reverse charge mechanism*, which is touched on in Box 2. We agree that this is not standard practice, but it was not presented as such and was only listed as one out of many tactics used to combat fraud.
The text acknowledges the presence of four types of MTIC fraud but groups the more sophisticated contra-trader and cross-invoicer schemes together. By grouping these schemes, the analysis may overlook the unique characteristics and challenges associated with each type of fraud.	While we acknowledge the unique characteristics of both of these types of schemes, they do not make a difference from the point of view of specific estimation methods. When using macro-level data there is no differentiation between contra-trader and cross-invoicer schemes and an in-depth discussion of their characteristics was not the objective of this report.

[Chapter I][…] an accurate portrayal of carousel MTIC fraud within the EU's consumption-based VAT system, adhering to the principle of destination, should consider the comprehensive revenue loss experienced by Member State 2 as follows. In essence, with a standard carousel, the revenue loss experienced by MS2 encompasses not only the unremitted VAT (=20) from the Missing Trader company B but also the refund (=20) owed to company C (the exporter). Consequently, MS2 incurs a total loss of 40 in revenue (twice the amount indicated in your cases, even without speculating on unknown price advantage).	In a situation where all the companies are following their tax obligations, the entire transaction chain (Company A → Company B → Company C → Company A) (assuming no margins/value-added) would be economically neutral in terms of taxation – this is because, in this scenario, the goods are only passing through Member State 2, to be exported further. In the case of carousel MTIC fraud, and when correctly interpreting the EU tax rules, the transaction would go as follows: Company A in Member State 1 exports the goods to Company B (Member State 2) for the price of EUR 100 (0% VAT). Company B then sells domestically to Company C, charging the net price + VAT (EUR 120) and is obligated to remit it to MS2. As we know, it does not do so and instead disappears with the VAT owed. Company C then exports back to Company A at a 0% VAT rate and is refunded the VAT it paid earlier by MS2. Overall, the only loss that MS2 incurs is due to the VAT unremitted by the missing trader, as in the case of further export it would still be refunding the VAT paid by Company C. As a side note, during our literature review we identified a similar, as suggested by the Reviewer, understanding presented in a briefing on MTIC fraud. Anyhow, we believe that we could not speak of double revenue loss in the case of a simple MTIC fraud or a single sequence of carousel fraud.
[…] the chapter lacks a clear structure, making it difficult to follow the flow of information. It jumps between different types of fraud without providing a cohesive narrative.	In response to this comment, we added additional content to better link the different types of fraud described in section I.a.
[…] it would benefit from more concrete case studies or real-life examples to illustrate the concepts and make them useful for a general taxonomy of the issue. What it is now reported in BOX 1 is not satisfactory/accurate.	The goal of the case studies presented in Box 1 was illustrating the level of complexity that some of these schemes can reach, rather than providing an exhaustive list of examples. Could you specify what kind of examples or cases you think are missing or would be beneficial to include?

[...] while the chapter highlights the complexities of MTIC fraud, it lacks a comprehensive exploration of potential solutions or best practices for addressing the issue. I found a limited discussion of countermeasures as the use of reverse charge (which is an exception, see above) which distort the analysis. [...] it would be helpful to include a discussion of the challenges faced by tax administrations and any ongoing efforts to address MTIC fraud. [...] The chapter could benefit from discussing the importance of international cooperation and coordination between tax authorities in combating MTIC fraud, as well as any existing initiatives or agreements in place, including properly the application of the reverse charge to which all the report is referred without considering that it is only a special case and only temporary allowed in the period of interest of the report. [...] while the box acknowledges the limitations of current measures, it does not delve deeply into assessing their overall effectiveness which is paramount also for evaluation. It mentions that new measures may cause fraud to shift, but it does not provide an in-depth analysis of the extent to which the measures have successfully reduced MTIC fraud or closed loopholes. The box does not compare the effectiveness of different anti-fraud measures nor discuss their relative strengths and weaknesses. While the box mentions the importance of coordinated actions among Member States to combat MTIC fraud, it does not thoroughly explore the challenges and limitations associated with achieving such coordination (which, by the way, also affect the possibility of estimation of the fraud). It could have discussed barriers to coordination, such as differences in legal systems, varying priorities among Member States, or bureaucratic hurdles. The box mentions various measures and procedures aimed at combating MTIC fraud, but it does not extensively address the enforcement aspect. While it briefly touches on penalties and due diligence requirements, it does not thoroughly analyze the enforcement mechanisms in place or the challenges faced by tax authorities in effectively enforcing anti-fraud measures. A more comprehensive examination of enforcement strategies, including resources, cooperation between tax authorities, and the effectiveness of penalties, would have provided a more robust assessment.	The elements mentioned by the reviewer are beyond the scope of the report. The measures introduced to fight fraud are discussed only to provide background and to the extent their introduction could be useful for the estimation of the fraud. The exploration of potential solutions is limited because Chapter I is focused, as the title would suggest, on mapping fraud pathways, rather than an in-depth discussion of best practices or challenges faced by tax administrations. This is why we chose to limit this discussion to Box 2, which considers not only the domestic reverse charge mechanism but also eight other measures, including those involving coordination at the EU level (the VIES, regulation on administrative cooperation). As for the remaining mentions of reverse charge, this is the mechanism which has been applied in the single market since 1993, see the first comment. It would be helpful to us if you could provide examples on how the analysis is distorted. We agree that this is only a limited exploration of the topic which could be expanded, but we fear that by including a lengthier discussion of international cooperation and best practices we would introduce confusion and decrease the readability of the whole chapter, which is entirely focused on the mapping and comparison on fraud pathways.

Finally, the text briefly mentions the European Public Prosecutor's Office (EPPO) and cross-border exchange of information, but it does not delve into the broader international cooperation efforts to combat MTIC fraud. […] and discussing international collaborations, information-sharing agreements, or best practices from other jurisdictions would have enhanced the analysis and highlighted potential areas for improvement	
[…] inadequate consideration of evolving fraud tactics. […] Failure to account for emerging fraud patterns and changing tactics can undermine the effectiveness of the proposed scenarios in identifying and preventing new forms of fraud and might affect the possibility of evaluating the phenomenon.	We believe that this is beyond the scope of this report. While MTIC fraud is characterized by dynamic changes in response to the external factors we listed in this chapter, the choice of methodology was informed by certain stable characteristics shared by all MTIC fraud schemes, discussed in Table 1.
[…] the chapter could benefit from a concise conclusion that summarizes the key points, reiterates the main challenges posed by MTIC fraud, and highlights potential avenues for future research and action.	Thank you for this comment, we added a concluding paragraph as suggested, with the exception of avenues for future research, which we believe did not belong in this and other chapters.
[Chapter II] […] lack of clarity in organization, repetition and redundancy. The text jumps between different types of sources and review processes without providing a clear structure. This makes it difficult to follow the flow of information and understand the purpose of each section. The text often repeats information or rephrases it without adding new insights. This leads to unnecessary length and makes it harder to extract the main points.	Thank you for this comment, the text was restructured to improve readability and repetitions were deleted.
[…] The text mainly describes the sources reviewed and their classifications without critically evaluating their strengths and weaknesses. It does not assess the reliability or validity of the sources or their methodologies, which is crucial also in policy oriented research. Moreover, it provides a list of sources and their classifications but fails to synthesize the findings or draw meaningful conclusions. It does not offer a comprehensive overview or analysis of the literature on MTIC fraud, which limits its usefulness for a policy-oriented report.	A discussion of the limitations within each “category” of approaches was included in the chapter. For better clarity the chapter was now restructured and includes a clearly marked section with this discussion and with conclusions.

It is acknowledged the scarcity of literature on estimating the scale of MTIC fraud, but the chapter does not explore alternative approaches or sources of data. It does not discuss potential solutions or recommendations for addressing the gaps in the literature.	Could you give examples of these alternative approaches/sources of data? We were unable to find any further literature that was relevant to the issue at hand and any suggestions would be much appreciated. In addition, could you also clarify what is meant by the discussion of "potential solutions or recommendations" for addressing the gaps? It is our core objective that our study will make a contribution to closing this gap.
The text briefly mentions input from reviewers and the Fiscalis Project Group but does not provide details about their involvement or the process of incorporating their suggestions. This raises questions about the transparency and inclusivity of the report's development. […] does not discuss the potential limitations or biases associated with relying heavily on unpublished materials.	Could you clarify what you mean by "providing details about their involvement", namely what information is still missing? Where appropriate the text underlines things which were the result of a reviewer's suggestion, and a complete table with comments from reviewers is provided in Appendix E. Furthermore, the Fiscalis Project Group was not involved in this project beyond providing one work-in-progress report and no greater involvement was suggested anywhere in the text. The report itself was simply included in the literature review, as mentioned here: *"The study team worked in close communication with the Fiscalis Project Group and received a work-in-progress report from them, which was taken into account in the review."* With regards to unpublished materials – we acknowledge that they have many limitations, which is why only one such report (mentioned above) was included. Considering that this was one out of over 70 sources used, we do not agree that the final report's reliance on such materials was "heavy" (and thus warranting a discussion).
I found inadequate the explanation of selection criteria and limitations. […] the chapter […] does not clearly specify the criteria used for determining relevance. […] The chapter does not provide a clear description of the methodology used for the review of literature. It does not specify how the sources were logically identified, screened, and selected for analysis.	The criteria for the selection and grouping of the papers were outlined in the paragraph directly preceding the table presenting the number of papers in each category (Table 2), which reads: *"Of primary relevance are empirical studies presenting own MTIC gap estimates, summary documents (such as reports from the European Commission), and literature presenting methodologies for estimating the*

	MTIC gap. Literature of secondary relevance includes studies presenting own MTIC gap estimates based on extrapolating from existing estimates or experts' qualitative assessments, and work describing the methodologies for detecting MTIC fraud. The remaining literature, which could not be assigned to either of the previous categories, was classified as "Other" (e.g., papers discussing the impacts of MTIC fraud or mechanisms of prevention), "Methodological" (concerning methodologies under consideration by the Team, rather than MTIC itself), or excluded from further analysis due to lack of relevance. There was no common theme running through the papers in this last category, but some examples included papers that were concerned only with the shadow economy or with VAT system design in general." Could you please clarify what you think is missing here? With regards to the methodology, the initial reports contained a more detailed description of the review process, which included information on how the papers were searched (where, with what keywords, in what manner). We chose not to include this description in order to not fatigue the reader and shift focus to the later chapters of the report, which are its key part.
While the text acknowledges the limitations of existing estimates and methodologies, it does not thoroughly discuss the implications of these limitations for policy-making or the reliability of the findings.	That was outside the scope of this study.
[Chapter III] The text states that "most Member States take active measures," but it does not specify what these measures are. Additionally, the phrase "until now" implies that the situation may have changed, but there is no further elaboration on the current state of affairs.	Chapter III is based on responses to the questionnaire sent out to Member States' administrations, therefore the information gathered is limited by the questions asked. In the report we wrote that conducting MTIC fraud detection on an annual basis *suggests* that active measures are taken, rather than asserting that that is indeed the case. Given that we did not include a question specifically regarding measures taken to combat fraud, we are not able to answer such a question. Of course, ideally the questionnaire would have been much longer and contained more questions, however this would likely negatively impact the response rate. We therefore focused on questions most relevant to this study (that is, connected to MTIC fraud estimation methodologies and establishing

	the data sources that could be used). We also note that the contents of the survey were not commented on in your previous review. What was meant by “until now” was “at the time of conducting the survey”. We agree that this is misleading and corrected this phrasing.
The information given is incomplete. The text mentions that 13 Member States conduct MTIC fraud detection on an annual basis, but it does not provide information about the remaining 6 Member States. Furthermore, it states that only six administrations have estimated the revenue losses due to fraud, but it does not properly mention the administrations surveyed.	The survey only asked whether the Member State in question conducted the estimation, and if so – was it on an annual or irregular basis. The answers to this question are presented in the graph and text. To the best of our knowledge, including a question probing this issue further was not suggested in the earlier review. The administrations surveyed are listed in the Appendix, they are not identified in the text due to confidentiality issues.
Figure 8 and Figure 9 are not properly explained in their content.	Could you expand on what remains unexplained? Currently the contents are explained in the paragraph preceding Figure 8, which summarizes the responses.
The chapter mentions the most and least common fraud detection and estimation methodologies, but it does not provide any benchmark or comparison with other types of fraud detection methods. […] The text presents information about the detection and estimation methodologies used by Member States but fails to discuss the implications of these findings. For example, it does not address nor mention the potential impact on the effectiveness of combating MTIC fraud or whether certain methodologies are more reliable than others. […] Various fraud detection and estimation methodologies are mentioned without providing detailed explanations of how these methods work or their strengths and limitations. This lack of information makes it challenging to assess the reliability and effectiveness of each approach.	This chapter is devoted solely to a discussion of Member States’ experience with MTIC fraud detection and estimation (specifically: whether or not they carry it out), based on responses to the questionnaire. Therefore, it serves more as an assessment of the current state of affairs. The comparison and reliability of those methods are discussed in chapters VII and VIII (and also includes administrations’ views on accuracy). However, we agree that this was not clearly stated in the text and may be confusing, so we added a paragraph at the beginning of Chapter III which clarifies this and tells the reader where this can be found. We chose to structure the report this way because we wanted to assess and compare the methods only after presenting the data considerations and assessment framework, and also because we wanted to avoid repetition in the report, which you noted in an earlier comment.
The text mentions that "two [administrations] do not [undertake the task of estimating revenue losses] on an annual basis" without providing any explanation for why this is the case. It would be useful to include reasons or factors contributing to the irregularity in these estimations.	The survey did not include questions regarding the reasons for which an administration did not carry out the estimation. To the best of our knowledge, including this question was not suggested in the previous review. Furthermore, it is possible that any answers to this question would be

	speculation, as the questionnaire was completed by individual representatives of said administrations. However, the most common reasons (the time and costs of such estimations) are mentioned multiple times in the text.
The text briefly mentions EU-wide information exchange between national tax authorities and the use of the Eurofisc network and TNA tool. However, it does not elaborate on the effectiveness or challenges related to these collaborative efforts.	That is outside of the scope of this study, which aims at selecting the methodological approach to MTIC gap estimation rather than to assess measures combatting the fraud.
The chapter mentions the number of Member States conducting MTIC fraud detection and estimation but does not provide any quantitative information regarding the scale or magnitude of the fraud. It would be helpful also in this chapter to include some statistics or estimates related to revenue losses or the overall impact of MTIC fraud.	We provided several estimates (EU-wide and for specific countries) in Table 3. There are almost no available estimates for MTIC fraud produced by individual Member States. In our survey only 4 Member State said they conduct such estimations at all and only 2 of them provided said estimates – differing in terms of years covered. We could provide those two, but they would be for different years, and we are not sure to what extent estimates for just 2 Member State would demonstrate the overall impact of MTIC fraud. This lack of estimates is exactly the motivation for the study.
The text mentions that only one Member State employs econometric modeling, but it does not provide any details on how this method is used or its potential advantages. Further elaboration on this approach and its applicability in detecting and estimating MTIC fraud would be beneficial.	Unfortunately, no further information on this method was shared with us.
The text does not discuss the legal frameworks or policy implications related to MTIC fraud detection and estimation. Including information on relevant EU directives, national legislation, or policy initiatives aimed at combating MTIC fraud would provide a more comprehensive analysis.	This is outside of the scope of this study, which aims at selecting the methodological approach to MTIC gap estimation.
The text concludes abruptly without summarizing the main findings or providing a clear takeaway. A well-crafted conclusion should summarize the key points, highlight significant insights, and potentially offer suggestions for future actions or research.	We did not write a summary because, in our opinion, this chapter offers clear takeaways along the way, which seems sufficient considering the text itself is only half a page long.
[Chapter IV] The chapter could benefit from clearer organization and structure. The information is presented in a somewhat scattered manner, making it difficult to follow the main points and arguments. Reorganizing the content and using subheadings to clearly separate different topics would enhance readability.	Could you expand on how you think the content should be reorganised or give examples of the lack of structure? Without this information it is difficult for us to properly address this comment, as at the moment this chapter is already divided into subchapters based on type of data.

While the chapter mentions various factors that contribute to discrepancies in mirror statistics, […] it lacks sufficient explanation and analysis of each factor.	An analysis of the factors contributing to mirror statistic discrepancies was not an objective of this study. This section served the role of making the reader aware that there is a lot of noise in this type of data and therefore not all of the discrepancies can be attributed to fraud. Even so, in many cases we fail to see the need to explain or analyse a factor in the first place, as most are quite self-explanatory (e.g., distortions caused by the application of exchange rates).
While the chapter provides a list of available data sources and their potential for estimating the scale of MTIC fraud, it does not include evidence to support the claims made.	Could you provide an example of claims which remain unsupported?
The subsection on data available for tax administrations presents an unsatisfactory summary of the responses received from EU Member States. Providing more context and details about the survey methodology would improve the transparency and reliability of the findings.	This comment is unclear – the subsection summarizes all the information on data availability that was collected via the questionnaire, no more and no less was collected. The questionnaire was sent to Member States' tax administrations to be completed (as is mentioned in the text of this section) and the questionnaire itself is enclosed in the report. We therefore fail to see the transparency issue you called attention to.
I noticed a lack of discussion on data quality and reliability: The chapter does not thoroughly address the quality and reliability of the available data sources. It is important to consider potential issues such as data accuracy, completeness, timeliness, and consistency across different sources.	The relevant discussion was extended. This is discussed in the assessment chapter to the extent we could judge on the quality of datasets that are not at our disposal.
The chapter primarily focuses on individual data sources without exploring the potential benefits of integrating multiple datasets. Data integration, combining information from various sources, can provide a more comprehensive view of MTIC fraud by capturing different aspects and cross-referencing information. Discussing the advantages and challenges of data integration would enhance the chapter's analysis. The methods or processes for validating and verifying the data obtained from different sources are not extensively discussed. Given the potential for discrepancies and inaccuracies in the data, it is important to consider validation techniques, such as data reconciliation, outlier detection, and cross-validation with other reliable sources. […] While the chapter mentions certain data gaps, such as the exclusion of specific categories of goods and services from Intrastat or the limited detail in trade in services	The chapter focuses on individual data sources because it is meant to serve as an introduction that demonstrates their usefulness, characteristics, and limitations. For this same reason it is not concerned with exploring solutions to the gaps present. Meanwhile, alternative data sources have been described at length in Chapter VII and the use of integrated data is a feature of most of the proposed methodological scenarios, as they use a combination of approaches, each drawing on different datasets (listed in Table 9).

statistics, it does not explore potential solutions or alternative data sources to address these gaps.	
The chapter briefly mentions the possibility of using hybrid estimation strategies and constructing mirror statistics, but does not delve into advanced analytics techniques that can be employed to detect patterns or anomalies indicative of MTIC fraud.	This chapter is devoted to data sources, not estimation strategies. We believe that this was a good decision to separate discussion on data and methods.
One additional aspect to consider is the scalability and sustainability of the proposed data considerations and methodologies. Considering the volume and complexity of trade transactions, it is important to assess whether the proposed data sources and methodologies can be applied consistently and effectively at a larger scale.	We assess the complexity and proxy the cost of implementation of every methodological scenario.
[…] the chapter could benefit from discussing the potential limitations and challenges associated with long-term sustainability. Changes in reporting requirements, data collection methods, or technological advancements can impact the availability and relevance of certain data sources over time. It would be valuable to address these potential challenges and suggest strategies for adapting data considerations and methodologies to evolving circumstances.	We believe that this beyond the scope of the study and what could be accurately projected.
The chapter concludes abruptly without summarizing the main findings or providing recommendations for policymakers or tax administrations. It would be helpful to offer key takeaways, highlight the most effective data sources or methodologies, and suggest areas for further research or improvement.	A comparison of methodologies is presented in Chapter IX (titled “Comparison”). With regards to recommendations for tax administrations and suggestions of areas for further research, this was outside the scope of the study.
[…] the chapter could better explore the potential for international collaboration and data sharing among EU Member States or other relevant entities.	That is outside of the scope of this study, which aims at selecting the methodological approach to MTIC gap estimation rather than to assess measures combatting the fraud.
[Chapter V] […] the chapter fails to provide a critical evaluation of the methodologies discussed. While it mentions different approaches, it does not assess their strengths and weaknesses, potential limitations, or areas for improvement.	Strengths and weaknesses assessed later, along with a discussion of how individual limitations might be addressed. Discussing areas for improvement of specific scenarios beyond perfecting them for the purpose of developing a common methodology was not the objective of the study.
It briefly mentions some statistical methods like k-means clustering and the Heckman procedure, but it does not provide sufficient information on how these methods work or why they are relevant in the context of MTIC fraud estimation. […] (in the current version, the additional information of boxes and table is not satisfactory).	We received conflicting comments that the discussion on the technical details of the methods was too extensive, rather than insufficient. We believe that the text (after some amendments) explains well how these methodologies would be used and we refer to other sources for detailed methodological considerations.

[…] briefly mentions the distinction between top-down and bottom-up methodologies but does not provide a thorough comparison of their advantages and disadvantages. A more comprehensive analysis should have included a detailed comparison of the two approaches, highlighting their respective strengths and weaknesses, and discussing when each approach is more suitable.	We believe that it is not the distinction between top-down and bottoms-up that explains the results of our assessment. Anyhow, we added additional content and believe that further generalisation could only be misleading.
The issue of data quality […] is not adequately addressed. The chapter briefly mentions controlling for "noise" in the data but does not elaborate on the potential challenges related to data accuracy, reliability, or completeness. […] lack of discussion on the limitations and potential biases inherent in the data used for estimating MTIC fraud: the data used for estimation may suffer from underreporting or misclassification of transactions. […] the potential strategies or adjustments to mitigate them are not discussed.	The relevant discussion was extended.
[…] does not address the issue of evolving fraud schemes and the adaptability of the estimation methodologies to detect new and emerging patterns of MTIC fraud. […] A discussion on the adaptability and robustness of the methodologies in the face of evolving fraud schemes would have been valuable.	This chapter was not intended to provide the assessment of methodological approaches but rather their classification and inter-linkages between different methods.
The chapter does not adequately address the issue of validation for the estimation methodologies discussed, nor it mentions the potential for evolving techniques and technologies in detecting and estimating MTIC fraud. It would have been beneficial to discuss the potential integration of emerging techniques into the estimation methodologies.	Could you please clarify what you understand by “the issue of validation for the estimation methodologies” and “potential integration of emerging techniques into the estimation methodologies”? This chapter provides classification of methodologies that form the methodological scenarios assessed by this report.
The cost-benefit analysis of the different estimation methodologies is not properly included. Assessing the costs associated with implementing and maintaining these methodologies in comparison to the potential benefits of reducing MTIC fraud would have provided a comprehensive perspective.	The estimated costs associated with the initial estimation and updates under each methodological scenario are presented in Chapter VIII, in tables 15 and 16 respectively. A comparison of each scenario using a Multi-Criteria Analysis approach is presented in the Assessment table (Table 17 & 18) in the same chapter. This “benefits” of each scenario relate to aspects of the resultant estimates, such as their accuracy or granularity. The criteria used draw on the robustness criteria outlined in the ToR. The potential benefits of reducing MTIC fraud have not been included because, first of all, they can only be quantified based on largely non-existent estimates of its scale (a gap which the methodology we’re developing is meant to address) and second of all, MTIC gap estimation itself will not directly lead to the gap closing or to increased detection of individual cases of fraud.

	The assessment of the impact of MITC gap monitoring on the potential reduction of its scale is beyond the scope of this report.
The chapter does not delve into the importance of international collaboration and the potential challenges in harmonizing methodologies and sharing data across different jurisdictions. Including a discussion on international cooperation would have enhanced the chapter's relevance in the context of combating MTIC fraud at a global level.	That was outside the scope of this study, please find an excerpt from the ToR outlining the study objectives at the end of this document.
[Chapter VI] […] the chapter starts by discussing the evaluation framework and alternative scenarios, but the transition between these two topics is not clear. The information presented could be organized more coherently to improve understanding.	We believe that at this page it shall already be clear to the reader that the objective of the report is to assess and compare methodological scenarios using the framework that is discusses in this chapter.
The chapter briefly mentions "the principles listed and discussed in Box 4," but it does not provide a clear explanation in Box 4.	Could you give an example of what is unclear or not sufficiently explained?
The terms "semi-quantitative" and "semi-qualitative" assessment are used to describe the evaluation process without clear definitions or explanations.	Added an explanation of the semi-qualitative assessment and referenced Table 7, which appears three pages later and provides detailed information on the modes of assessment.
MCA […] subjectivity introduces a potential weakness in the assessment process […]. The criteria for MCA are assigned weights based on priorities discussed and scored with the assistance of the Commission, but specific details on how these weights are determined or how the scoring process is conducted are not provided. […] The chapter does not adequately address potential biases in the assessment process. […] such as the selection of evaluation criteria, the weighting of criteria, or the involvement of specific experts or stakeholders. It is important to explicitly acknowledge and mitigate these biases […]	The subjectivity of MCA was acknowledged but it is unavoidable given that the study consists of multiple objectives and different stakeholders will have different views on the order of priorities. This issue was addressed in Box 4 and the use of a Monte Carlo simulation was suggested. Please could you provide an example of details that are still missing from the discussion? The choice of evaluation criteria was based on the objectives listed in the ToR.
The chapter briefly mentions some preselection criteria […] However, it does not elaborate on how these criteria are evaluated or provide specific details about the thresholds or standards used for preselection. […] it is mentioned that preselected methodologies are scored using parameters described in Table 7, but a thorough explanation of how the scoring process works is not provided. The lack of clarity regarding the assessment methodology makes it difficult to evaluate the validity and reliability of the scores assigned to the methodologies.	Table 7 provides all the information you listed (method of evaluation, thresholds), column (3) in that table specifies whether or not a given criterion was considered in the preselection.

The chapter introduces a list of criteria and sub-criteria for evaluation without providing a clear justification for their inclusion. The chapter focuses primarily on internal criteria related to the quality and availability of information, costs, and risks. However, it does not sufficiently address the potential impact of external factors on the assessment process. Factors such as changes in regulatory environments, technological advancements, or emerging fraud schemes could significantly influence the suitability of different methodologies. Ignoring these external factors limits the comprehensiveness of the assessment.	We believe that the rationale for the list of the included criteria is comprehensive. Criteria cannot be "internal" – criterion is not a factor but a dimension/function that is used in the assessment. The assessment process – assigning values to different criteria takes into account both factors beyond and outside of the modeller's control.
The chapter briefly mentions the availability of data as a criterion for method selection. However, it does not delve into the potential limitations and challenges associated with data collection and availability. Issues such as data accuracy, timeliness, and consistency across Member States could significantly impact the reliability of the chosen methodologies.	There is no such criterion used. Data availability impacts some scores (e.g., completeness) but cannot be a criterion per se. In other words, it is not an objective to use that the data. The objective is to derive accurate and complete (please also see other criteria), which obviously requires accurate and complete data.
The chapter briefly mentions the feasibility of producing estimates for at least 14 Member States and a specific time period. However, it does not address the scalability of the chosen methodologies beyond this limited scope.	We see no reason why scalability should be a criterion in the comparison framework. Scalability is related to the effort and complexity, which are included in the assessment.
The chapter mentions that two criteria were assigned categorical scales without providing further details on the specific scale values or the rationale behind them. […] Providing clearer guidance on the categorical scales would improve the transparency and consistency of the assessment process.	Additional explanations were added. For details please see Table 7. This is an example – description of scoring for a specific criterion: "Ability to link the value of the VAT fraud tax gap to specific drivers/types of fraud": 0: No breakdown 0.5: Possibility of breakdown by type of irregularities or type of taxpayers 1: Possibility of breakdown by type of irregularities and type of taxpayers
The chapter focuses on the evaluation of different methodologies but does not discuss the practicality of implementing these methodologies in real-world scenarios. Factors such as resource requirements, data collection processes, and the capacity of Member State administrations to adopt and implement the chosen methodologies should be considered. Ignoring practical implementation considerations may result in the selection of methodologies that are challenging to implement effectively.	Could you please explain what you mean by "real-world scenarios"? The objective of the study is to select the best approach to implementing the calculation under Phase II of the project.

The potential of innovative or emerging techniques for MTIC gap estimation should be explored. Given the dynamic nature of tax fraud schemes, it is crucial to consider novel approaches and technologies that may offer improved accuracy and effectiveness in detecting and estimating the MTIC gap.	We propose novel approaches but cannot assess anything that has not been invented or we cannot think of. These are "unknown unknowns".
The chapter lacks a detailed discussion on uncertainty analysis and sensitivity testing [...]. Including a thorough analysis of uncertainty and sensitivity would provide a more comprehensive evaluation of the methodologies.	Sensitivity check results and discussion included in the final version of the report, section IX.b. (comparison chapter).
A mechanism for continuous improvement or ongoing monitoring of the chosen methodology should be considered. Given the evolving nature of tax fraud and changing regulatory environments, it is important to establish a feedback loop to evaluate the effectiveness of the chosen methodology and make necessary adjustments over time.	We concur but this is beyond the scope of Phase I of this study. It could only be carried out when the selected scenario is implemented.
The chapter mentions that the assessment relies on information gathered from the literature and experience of Member State administrations. However, it does not address (not even in other chapters) the potential biases or limitations of these sources.	The limitations of these sources were pointed out in chapters II and VIII (section on perceived accuracy). We now reorganised these sections to help these comments stand out. However, it must be kept in mind that the sole objective of these sections was the mapping of methods used up to date (the time of writing the report) and any further discussion included was already outside the scope of this study.
The chapter [...] does not provide sufficient detail on how the CBA is conducted or how the costs and benefits of different methodologies are assessed. A more thorough explanation of the CBA methodology and its application would strengthen the analysis of costs and benefits associated with the chosen methodologies.	As noted in the chapter, we used a combination of CBA and MCA, due to the nature of some of the assessment criteria. Upon reviewing the chapter once more, we are not sure what further explanation can be provided beyond what is presented in Table 7 (the Assessment column). Could you please specify what kind of details you think are missing from what is currently presented?
The potential impact of technological advancements on MTIC gap estimation methodologies should be adequately address. Technologies such as artificial intelligence, machine learning, or data analytics may offer new opportunities for improving the accuracy and efficiency of estimation.	The objective of the study is to select the methodology that could be maintained in the mid-term rather that assess the future of the MTIC estimation in the long-term.
The chapter does not explicitly address the dynamic nature of MTIC fraud patterns and the issue of the long-term sustainability of the chosen methodology. [...] Incorporating a mechanism for monitoring and responding to changing fraud patterns would ensure that the chosen methodology remains relevant and effective over the long term.	Again, the scope of the study was limited to identifying a methodology which would be workable and fulfill the set out objectives in the present day. Moreover, the dynamic nature of MTIC fraud was discussed at length in earlier chapters, where we also outlined certain stable characteristics shared by all MTIC fraud schemes. It is those characteristics that informed the choice of methodology.

The chapter does not sufficiently address the limitations and assumptions associated with the evaluated methodologies.	This chapter provides the framework. The methodological scenarios are assessed in Chapter VIII.
The chapter does not outline a process for continuous evaluation and revision of the chosen methodology. […] The inclusion of benchmarking or external validation as part of the assessment process should be considered as well as the interdependencies between the evaluation criteria should be thoroughly explored. Comparing the performance of the evaluated methodologies against established benchmarks or validating the results against independent data sources could provide additional credibility and confidence in the chosen methodology. Some criteria may have direct or indirect relationships with each other, and failing to account for these interdependencies could lead to suboptimal or inconsistent assessment results.	We concur but this is beyond the scope of Phase I of this study and beyond the description of the assessment framework. This concerns the operationalisation of the calculation and not the selection of the approach.
The chapter does not provide a clear explanation of the criteria used for final selection and the relative importance assigned to each criterion. Transparently defining the decision-making criteria and their relative weights would enhance the clarity and fairness of the assessment process.	Could you specify what information is in your opinion still missing? The criteria used for the final selection, along with their weights, floors, ceilings, and method of assessment are all listed in Table 7 and in the comparison tables in Chapter IX.
Finally, the chapter should thoroughly explore the potential trade-offs that may exist between different evaluation criteria. […] Providing a nuanced discussion of these trade-offs would help decision-makers understand the inherent compromises involved in selecting a methodology.	We added additional explanations on the inter-linkages. Yet, the study is designed (as explained in the introduction) to support the choice of the methodology by EC rather than MS' administrations that may have different objective behind the MTIC gap estimation.
[Chapter VII] The chapter lacks clarity in its presentation. It jumps between different sections and topics without a clear structure, making it difficult to follow the flow of the discussion. Some information is repeated multiple times, such as the mention of the importance of controlling for factors other than MTIC fraud in the analysis of trade data.	Could you please provide an example? In its current form, the first paragraph of the chapter outlines a clear structure that the chapter does not veer from (promising methods list – discussion of those meeting pre-selection requirements – grouping into scenarios) While we do mention controlling for noise several times in this chapter, because we discuss it with regards to several methods, I could not find the repetitions you speak of, concerning its importance. We only comment on whether or not a given method has this ability, which I find is relevant given that it is one of the main goals we set in that chapter. If there are any other repetitions you caught, could you give specific examples?
[…] chapter does not provide sufficient justification for the selection of certain methodological approaches over others. It mentions that certain methods have been	We believe that the entire content of this chapter is devoted to the justification for the selection of methodological approaches.

shortlisted or proposed for further evaluation, but it does not clearly explain the criteria or rationale behind these selections. […] does not compare the different proposed methodologies in terms of their strengths and weaknesses. It does not provide a critical analysis of the advantages and disadvantages of each approach, making it difficult to assess their relative merits. […] It is acknowledged that all the proposed indicators and methodologies have substantial limitations. However, the chapter does not provide a comprehensive discussion of these limitations or address the uncertainties associated with the estimates obtained through these methods. The reader is not informed about the potential margin of error or the reliability of the results.	We do not compare the approaches/analytical methods but the entire "scenarios". Why we do so in explained and repeated throughout the chapters.
It does not provide a comprehensive assessment of the strengths and weaknesses of each data source, which hinders the reader's understanding of their reliability and suitability for the task.	This chapter in neither devoted to data considerations nor the assessment of methodological approaches (please see relevant chapters IV and VIII).
It mentions the need for validation of the selected methods but does not elaborate on how this validation process should be conducted or what criteria should be used. It does not discuss the importance of independent verification or the potential limitations of validation approaches.	Guidance on how to implement each of the methodological scenarios is beyond the scope of this report.
[…] lack of empirical evidence, inadequate discussion of legal and ethical considerations, and insufficient discussion on validation methods.	We believe that this is beyond the scope of this chapter.
[…] limited consideration of alternative approaches: The chapter briefly mentions the potential use of structural equation modeling (SEM) and the Kalman filter […]. However, it does not thoroughly discuss the strengths, weaknesses, and applicability of these methods. The reader is left without a clear understanding of how these approaches differ from the others presented and why they are being considered (BOX's explanation is inadequate at the moment).	We believe that the elements mentioned by the Reviewer are present in the text. Some example (for SEM): *There are three major reasons for using latent variable models. First, the approach is parsimonious because these models can summarise information contained in many response variables using only a few latent variables. Secondly, when properly specified, a latent variable model can minimise the biasing effects of measurement errors on estimating treatment effects. This property often makes this approach more accurate than a traditional version of the same analysis. Thirdly, latent variable models investigate effects between primary conceptual variables rather than between any particular set*

	of ordinary response variables. Consequently, a latent variable model is often viewed as more theoretically appropriate than a simpler analysis with response variables only (Breusch, 2005, direct citation). *This use of structural equation modelling of the underground economy has its critics. According to Kirchgässner (2016) MIMIC is a confirmatory, rather than exploratory, statistical technique. It is not valid to conclude that a variable has been found as a statistically significant determinant of the shadow economy. As pointed out by Dybka et al. (2019), there is also a strand of research showing that the results of the MIMIC tend to be unstable. In addition, there is no economic theory to guide the specification, and the complexity of the estimation strategy is also often criticised. Moreover, the underground economy is not a latent or hypothetical quantity like intelligence; it is all too real, just difficult to measure (Breuch, 2005).*
[…] limited discussion of validation and robustness: The text mentions the need for validation of the selected methods but does not provide a thorough discussion of how the validity and robustness of the estimates will be ensured. It does not discuss the importance of sensitivity analysis or the need to test the methods on different datasets or time periods to assess their reliability. […] absence of sensitivity analysis […] which explores the impact of variations in key parameters and assumptions on the results, is crucial in assessing the robustness of the estimation methods. The chapter does not discuss its inclusion, which leaves the reliability and stability of the estimates unexplored.	Again, this is beyond the scope of this study (and scope of all the chapters that attracted similar comment).
[Chapter VIII] While the chapter discusses sampling error and its relation to sample size and standard deviation, it does not provide a comprehensive assessment of the statistical error. It does not discuss other sources of statistical error or their potential impacts on accuracy estimation. Factors such as the standard deviation of unexplained error and the share of variation explained by exogenous variables also affect the statistical error. Non-sampling errors, on the other hand, do not diminish with sample size and are often related to issues like non-response, inaccurate reporting, or violations of assumptions in statistical models. These issues should be better assessed.	Thorough assessment of all the components of the error bot statistical and non-statistical before the implementation of these methods is, unfortunately, not possible. For this reason, we implemented (using the training datasets) most promising approach to be able to develop this discussion.

[…] a comprehensive analysis of the potential sources of error and uncertainty in the proposed methods is missing. While the second part briefly mentions statistical and non-statistical error components, it does not provide a thorough examination of their impact on the accuracy of the results. Understanding and quantifying these sources of error is crucial for assessing the reliability and confidence in the estimated VAT compliance gap and fraud detection outcomes.	
One weakness that emerges in the first part is the lack of specificity and detail regarding the proposed methodological scenarios for MTIC fraud detection and estimation. While the use of risk analysis systems and machine learning techniques are mentioned, there is not clear explanations of how these methods are applied or the specific algorithms and models used. This lack of detail makes it difficult to assess the robustness and effectiveness of the proposed scenarios.	The details for the implementation of these methods by respective administrations are largely confidential. Even if it was not, adding another layer of detail (as variables used by different risk-analysis systems) would not support the objective of this report.
[…] limited discussion on the potential limitations and challenges of the proposed scenarios. While the text acknowledges that no method is foolproof and that each approach has its own strengths and weaknesses, the chapter does not delve into a comprehensive analysis of these limitations. […] While the chapter briefly mentions the limitations of individual indicators and scenarios, it would be helpful to provide a more comprehensive discussion of the specific weaknesses and potential drawbacks associated with each methodological scenario.	Limitations and challenges of the proposed scenarios are grasped by the quantification/assessment of each of the criteria included in the analysis.
A further weakness is the limited availability of data for assessing different methods. The text mentions that the available evidence was insufficient for comparing the results of different methods or model specifications implemented in the past, which hinders the ability to gain insights into the accuracy of different approaches. This limitation reduces the overall confidence in the proposed scenarios and their effectiveness in addressing MTIC fraud detection and estimation.	This is, unfortunately, beyond our control.
[…] the issue of data quality and integrity is not addressed adequately. The second part of the chapter acknowledges that the quality of data from the Intrastat International Trade dataset may not always be ideal and that significant data cleaning efforts may	Biases in the data that are unavailable to the study team are largely unknown, so such an assessment cannot be made. For all the available series, we thoroughly assess the data. Intrastat data have their dedicated sub-chapter (IV.a).

be required. However, it does not delve into the specific challenges and potential biases that may arise from working with imperfect data.	
As in previous chapters, there is again the limited consideration of the dynamic and evolving nature of MTIC fraud. […] proposed scenarios seem to rely on static approaches without accounting for the potential changes in fraud patterns over time.	See previous replies to this recurrent comment.
[…] lacks a comprehensive assessment of the scalability and generalizability of the proposed scenarios. While it discusses the methods and findings […], it does not provide insights into how these scenarios can be applied in a broader context or scaled up for use across different countries or regions.	See previous replies to this recurrent comment.
[…] texts provide insights into specific methods […]. However, there is a lack of comprehensive evaluation and comparison of alternative methods. A broader analysis and comparison of various techniques would enhance the understanding of their strengths and weaknesses, leading to more robust and reliable approaches.	Could you give an example of what is missing in the current evaluation?
[…] the assessment of complexity and costs appears to be based on averaged responses and assumptions, rather than a detailed analysis. The lack of comprehensive and specific information about the complexity and costs of implementation may hinder accurate planning and resource allocation for implementing the proposed scenarios.	We believe that this is beyond the scope and possibility of implementation.
Conduct a comprehensive cost-benefit analysis for each methodological scenario. Evaluate the potential benefits of implementing the scenarios in terms of increased revenue collection and improved fraud detection against the costs associated with data collection, analysis, and resource allocation. This analysis will help policymakers make informed decisions on the most cost-effective and impactful approaches.	The aim of the study is to develop a common methodology for MTIC gap estimation, not the detection of MTIC fraud. While having estimates which are more accurate and comparable across Member States would indirectly help in closing this gap, as it would allow Member States to closely monitor the impact of introduced anti-fraud measures (a benefit which is nevertheless difficult to quantify), it would not directly lead to increased revenue collection or improved fraud detection.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696 or

– by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

KP-05-24-169-EN-N

Publications Office
of the European Union

DOI 10.2778/25011
ISBN 978-92-68-13299-9