**IFC-Bank of Italy Workshop on "Data science in central banking: enhancing the access to and sharing of data"**

**17-19 October 2023**

# EMIR data for financial stability analysis and research[1]

Michele Leonardo Bianchi, Bianca Sorvillo, Dario Ruzzi, Federico Apicella, Luigi Abate, Leonardo Del Vecchio, Bank of Italy

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

# EMIR data for financial stability analysis and research

Bianchi, Michele Leonardo; Sorvillo, Bianca; Ruzzi, Dario; Apicella, Federico; Abate, Luigi; Del Vecchio, Leonardo[1]

## Abstract

This work aims to set best practices for the analysis of the novel and extensive European Market Infrastructure Regulation (EMIR) dataset. By comprising trade-level information on millions of derivative transactions every day, this dataset is particularly useful to supervisors, regulators and central banks for monitoring financial markets by means of real-time risk metrics. Data granularity, however, comes at the cost of high complexity compounded by the dual-reporting nature of EMIR (as both 'legs' of a transaction may be subject to reporting obligations). After highlighting the data quality and practical issues faced when collecting and using EMIR data, we propose a framework to move from raw to clean data. Moreover, we provide statistics on notional values, based on which we visualise interconnections among counterparties.

Keywords: EMIR, granular data, derivatives, financial stability

JEL classification: G18, G00

[1] Financial Stability Directorate, Bank of Italy. The authors thank G. Papadia, D. Zambuto and G. Turturiello for technical support. The views expressed are those of the authors only and do not necessarily reflect those of the Banca d'Italia.

# Contents

**EMIR data for financial stability analysis and research** ..... 1
Summary ..... 3
1. Introduction ..... 5
1.2 European Market Infrastructure Regulation (EMIR) ..... 5
1.3 Overview of existing research work on EMIR data ..... 7
2. Data collection process through trade repositories ..... 8
3. EMIR Data ..... 10
3.1 Data overview ..... 10
3.3 Data processing framework ..... 15
3.4 Dealing with collateral portfolios ..... 18
3.5 Comparison with other sources ..... 19
3.6 Market statistics ..... 21
3.7 Some EMIR network snapshots ..... 23
4. Conclusion ..... 25
References ..... 26
Appendix ..... 28
A.1 Reporting of centrally cleared trades ..... 28
A.2 Data processing and confidentiality issues ..... 29
A.3 EMIR introspection for the adjustment of GLEIF data ..... 29

## Summary

EMIR data can be used in many areas relevant to the work of supervisors, regulators and central banks, which include financial and market supervision, financial stability, research and statistics. In such areas EMIR data are useful to build real-time risk metrics and to visualise and capture interconnections between counterparties, hence supporting authorities' monitoring of financial markets.

The main purpose of this work is threefold: (i) to give a broad picture of the EMIR data, (ii) to highlight some practical issues associated with them, and (iii) to present some methods and solutions that analysts and researchers working in competent authorities may adopt when using these data. The issues in questions are related to data quality, to the two-sided reporting mechanism and to the quantitative information on margins, which are reported at the collateral portfolio level rather than at the contract level. The analyses presented in this paper are based on EMIR data available to Banca d'Italia, which is among the competent authorities with access to them.

Each day the EMIR dataset increases by millions of entries and a big data platform is needed to collect, store, process and analyse the data. Unlike other datasets available to competent authorities, EMIR data are not directly collected by them: data come from four trade repositories (TR) and ESMA provides a single access point to these data. Another key characteristic is that EMIR data are collected on a daily basis: this means that today one has the possibility to examine the derivative positions with reference to normally two trading days ago. This is of practical relevance for financial stability monitoring tasks.

After a brief description of the main aspects of the regulation on reporting and of the data collection process, we provide an overview of existing research works on EMIR data. Then, we move our focus to the data with the aim (1) to have an overview of the available information; (2) to define a framework to process the raw dataset in order to obtain a *proper* dataset for the analysis; (3) to provide some market statistics; (4) to visualise interconnections among counterparties.

More in detail, we first explore the qualitative information in the dataset. This first step is focused on the information on the counterparties, including their nationality, and on the characteristics of the contracts, including the contract type, the asset class of the underlying, the currencies of the legs, the execution venue and the information on clearing and collateralisation.

Then, we analyse one of the most important quantitative fields, which is the notional value. By leveraging on this value and on some other additional fields, we define a data cleaning framework for the overall data set. Additionally, two-sided reported trades are deduplicated[2] according to a pecking-order criterion centred around reporting reliability to avoid double counting of notional values. We then describe how to deal with collateral portfolios[3] to extract initial and variation margins.

The data processing framework allows us to explore the quantitative information related to the notional value, to compare the data with other sources (e.g. FINREP and the supervisory and statistical reports), to provide some derivative market statistics evaluated on the outstanding contracts, and to visualise interconnections between market participants.

Transactions where one of the counterparties is an Italian resident or where the underlying is Italian are only a small part of the EMIR dataset Banca d'Italia has access to. Most of them, in terms of the number of submissions, are reported by Dutch, Cypriot, and German counterparties. In terms of notional amounts, German entities are the major reporting counterparties, before Dutch and Cypriot entities. In their role of reporting, Italian entities are below five per cent of the total notional of outstanding contracts.


[2] If both counterparties submit a report for the same trade, only one is selected on the basis of the deduplication algorithm.

[3] To minimise the costs of collateralisation, the management of the collateral related to derivatives contracts between two counterparties is usually performed by grouping contracts with the same characteristics or underlyings (i.e. collateral portfolios).

The main findings of our analysis conducted on the data between January 2021 and April 2023 are the following:

- there exist thousands of reporting counterparties (considered at an individual level), for some of which there are millions of reports;
- the combination of the cleaning[4] and deduplication processes applied to the raw dataset decreases significantly the number of records and trades to be included in the analyses;
- as far as Italian banking groups are concerned, the notional amounts reported in EMIR are close to those reported in supervisory (FINREP, harmonised at the EU level) and statistical reports;
- a substantial portion of the total notional value is associated with derivatives on interest rates, particularly swaps; equity derivatives show the second largest share of the overall notional value, while credit and currency derivatives provide only modest contributions and commodity derivatives represent the smallest fraction.

Finally, with a new EMIR reporting regime entering into force in April 2024, the number of fields will increase from around 100 to around 200, in order to foster the alignment with the best market practices and to improve the information content of the submitted trades.

[4] Intragroup transactions are discarded in this analysis.

# 1. Introduction

European Market Infrastructure Regulation (EMIR) data can be used in many areas relevant for supervisors, regulators and central banks. These areas include financial and market supervision, financial stability and statistics. The main purpose of this work is to help analysts and researchers who work in competent authorities and wish to deal with EMIR data by also shedding light on some practical issues they may face when using this dataset. For example, EMIR data can be used to improve the authorities' understanding of how financial markets really work, to build real-time risk metrics, to visualise and capture the interconnections between counterparties.

EMIR dataset comprises millions of entries – each representing a derivative transaction – every day. The dataset complexity has two dimensions: (1) the complex nature of derivative instruments; and (2) the huge amount of data. A big data platform is needed to collect, store, process and analyse these data.

Unlike other datasets available to competent authorities, EMIR data are not directly collected by them (see Section 2 on the data collection process). This means that there is no direct communication channel between reporting entities and most data users, something that instead is helpful when for instance double checks on submissions from reporting entities is needed. As we will see in the section on trade repositories (TRs), the journey of reports from the reporting entity to the analyst working with EMIR data is far from being linear and asking for an explanation of a single submission can be challenging (not to say impossible).

Another key characteristic is that EMIR data are collected on a daily basis. This marks an important difference with the typical supervisory data which are collected only on a monthly (or quarterly) basis (e.g. Banca d'Italia is informed of the exact composition of banks' bond portfolios of supervised entities with a monthly frequency). Accessing EMIR data allows to know the trade-by-trade activity on derivatives of the supervised entities. Additionally, EMIR has also information on derivatives of non-supervised entities (e.g. non-financial companies).

After having briefly described the main aspects of the regulation, Section 1.3 reviews the research work carried out so far on EMIR. Section 2 explains the data collection process, while Section 3.1 provides an overview of the main collected information and on the reporting entities. Data quality issues and the processing framework are discussed in Sections 3.2 and 3.3, respectively. In Section 3.4 we show how to deal with collateral portfolio codes to extract quantitative information on initial and variation margins. By focusing on Italian banks, Section 3.5 compares the notional amounts extracted from EMIR data to those reported in FINREP and in the supervisory and statistical reports. Section 3.6 discusses the main derivative market statistics. In Section 3.7 we show examples of networks built with EMIR data. The analyses presented in Section 3 are based on EMIR data available to Banca d'Italia as one of the relevant competent authorities. In the Appendix we show how centrally cleared trades should be reported, we describe how to access the big data platform, some confidentiality issues and how to extract some information on the group structure from EMIR reports.

## 1.2 European Market Infrastructure Regulation (EMIR)

With the aim of improving stability, transparency and efficiency in derivatives markets, on 16 August 2012, Regulation (EU) No 648/2012, better known as European Market Infrastructure Regulation (EMIR), entered into force. EMIR is based on the following principles: clearing, margining and reporting. It also provides a framework for authorisation and supervision of CCPs under the responsibility of the national competent authorities (NCAs), as well as registration (or recognition) and supervision of TRs through the European Securities and Markets Authority (ESMA).

Under EMIR, there are three main obligations on market counterparties that hold for any form of derivatives: (1) to report all derivative contracts to a TR recognised by ESMA; (2) to centrally clear some

OTC derivatives; (3) to apply risk mitigation techniques for OTC derivative contracts not cleared by a CCP, such as confirmation, daily mark-to-market, portfolio reconciliation and exchange of collateral.

To reduce regulatory burdens and compliance costs and to improve transparency and data quality, EMIR has been amended by Regulation (EU) No 834/2019, in force from 2019. The most significant changes concern the categorisation of counterparties for the purpose of applying the clearing obligation and the exemption, under some circumstances, from the reporting obligation of intragroup derivatives contracts. Regulation (EU) 2019/363 amends the fields to be reported to TRs and defines the reporting template currently in force. Among the various regulatory documents of the EMIR and the Securities Financing Transactions Regulation (SFTR) package, Regulation (EU) 2019/361 concerns the access to the data held in TRs by the competent authorities. Direct and immediate access to derivatives data is essential to allow authorities to fulfil their responsibilities and mandates (see also Section 4.1.3.2 in ESMA, 2023).

ESMA and the European Systemic Risk Board (ESRB) have access to the full EU-wide dataset. Banca d'Italia has access to those transactions where one of the counterparties is an Italian resident or where the underlying asset is Italian. Additionally, Regulation (EU) 2019/361 grants Banca d'Italia access to a broader perimeter including all transaction data:

- on derivatives in either of the following cases: (i) where the reference entity of the derivative is established within a Member State whose currency is the euro and falls under the supervisory responsibilities and mandates of Banca d'Italia; (ii) where the reference obligation is sovereign debt of a Member State whose currency is the euro (see article 1, paragraph 9);
- on derivatives concluded on trading venues or by CCPs and by counterparties that fall under responsibilities and mandates of Banca d'Italia when monitoring systemic risks to financial stability in the euro area (see article 1, paragraph 10).

The implementation of the regulation seems not homogeneous among different TRs. We detected some issues which resulted in either underreporting and possible overreporting of EMIR data from TRs to Banca d'Italia. We are interacting with TRs and ESMA to solve these issues (see also Section 4.1.3.2 in ESMA, 2023, where the EMIR regulatory access filtering is discussed).

For every derivative contract, both counterparties are required to submit details of who they are trading with and the details of the trade. The reporting is conducted at trade level.[5] The dual nature of reporting (i.e. by both counterparties of each contract) is for data quality purposes and reduces the incentive to omit trades.[6] Reports are submitted to a TR, where the data is aggregated and stored (see Section 2). There are over 100 fields in each report, although not all fields apply to each trade.[7] The information to be reported includes data on counterparties and data about the contract, and is expected to be the same for both counterparties (i.e. *common data*). Details about the characteristics of the contract are also requested, including qualitative and quantitative information on the contract itself, on its underlying and on the collateral (when the contract is collateralised). Additionally, for quantitative information on contract value, price and margins, mandatory reporting on a daily basis is requested up to the maturity or the early termination of the contract. Under EMIR, initial and variation margins are calculated and reported, based on the net positions resulting from a set of contracts (*collateral portfolio*). Margins must

5 It should be noted that is not possible to understand if the derivative is used for trading, hedging or other purposes. There are only two fields (i.e. "*trading capacity*" and *"directly linked to commercial activity or treasury financing"*) that may help to understand if the reporting counterparty has concluded the contract as principal on own account (on own behalf or behalf of a client) or as agent for the account of and on behalf of a client and, if the contract is directly linked to commercial or treasury financing activities. The quality of these two fields does not seem adequate for their use.

6 Firms can choose to delegate the reporting obligation to the other counterparty or a third-party. As usual, the delegating firm remains responsible for the accuracy of the reports.

7 See the Commission Implementing Regulation (Eu) 2019/363 of 13 December 2018.

be computed at the collateral portfolio level, rather than at the contract level. As discussed in Section 3.4, dealing with this reporting framework is challenging from a computational perspective.

A new reporting regime will enter into force in April 2024 with the aim of improving the clarity of the reporting framework.[8] This regulatory package defined in 2022 and known as EMIR Refit amends the details of the data to be reported to TRs by increasing the number of fields from around 100 to around 200 (see the regulatory and implementing technical standards 2022/1855 and 2022/1860).[9] It is to be noted that the major part of the increase in the number of fields is to align the reporting to the best market practices and to improve the information content of the submitted data.

## 1.3 Overview of existing research work on EMIR data

Empirical studies on granular derivatives data are relatively few, mainly due to limited data availability (see also the Annex in ESMA, 2023). Since 2014, EMIR started to require EU institutions to report their derivative transactions to TRs and data availability in Europe has largely improved since then. TR data are confidential and only ESMA, ESRB, European Central Bank (ECB) and NCAs have access to these transaction-by-transaction data. Additionally, a big data infrastructure is needed to store, manipulate and analyse this large amount of data. Confidentiality rules, together with technical complexities, slow down the development of a specific stream of research and papers exploring EMIR data have been published in peer-reviewed academic journals only recently.

The research on EMIR data can be divided into three main categories: (1) notes on data description (including on data quality issues); (2) studies on specific contract types or particular market events; (3) risk analyses based on network methodologies.

The description of EMIR data is not a simple task. After an overview of the EMIR dataset and of the data cleaning procedures, Abad et al. (2016) provide descriptive statistics of interest rate, FX, and CDS markets. Ascolese et al. (2017), van Lelyveld (2017) and Ullersma and van Lelyveld (2022) explore the potential and challenges of EMIR data for policymakers and regulators. Interesting papers on data cleaning methodology are those by Pérez-Duarte and Skrzypczynski (2019) and Lenoci and Letizia (2021).

Abbassi and Bräuning (2021) study German banks hedging strategies based on FX forward contracts. Numerous analyses have been conducted on credit default swap (CDS) contracts (see Kenny et al., 2016, Levels, 2018, Abraham, 2020, Pedro, 2020, Bianchi, 2021, Mazzacurati et al., 2021). Among CDS contracts, in fact, there are the single-name CDS that are simple to analyse, given that their information is more standardised[10] and the risk direction can be easily identified. Bellia et al. (2018) examine the clearing choice of counterparties stipulating CDS with underlying of Italy, France and Germany's sovereign debt. While Bias et al. (2020) shed light on the derivatives use by equity mutual funds, Kenny (2016) and Fiedor and Killeen (2021) focus on securitisation special purpose entities.

Some researchers investigate specific market events. Cielinskaet al. (2017) and Joseph and Vasios (2022) investigate the impact of the Swiss National Bank's decision to discontinue the floor of 1.20 Swiss francs per euro on the morning of 15 January 2015 on the foreign exchange derivatives market. Schroeder (2020) examine the underlying drivers of the flash crash in Pound Sterling vs US Dollar in October 2016 and the contribution of different market participants to the drying up of liquidity during that flash crash. Bouveret et al. (2022) explore Archegos positions and show that it is possible to track the steep increase in concentrated exposures that the family office undertook in February and March 2021.

8 See the ESMA Guidelines for reporting under EMIR of 14 December 2022 and the Commission Delegated Regulation (EU) 2022/1855.

9 See also ESMA (2022b).

10 Their underlying is usually identified by an International Securities Identification Number (ISIN) or a Legal Entity Identifier (LEI) code.

Some papers examine initial and variation margins (see ESRB, 2020, and Cominetta et al., 2019) also with a focus on risk management practices of central counterparties (Grothe, 2021). D'Errico and Roukny (2021) use bilateral obligations resulting from CDS contracts to explore the mechanics of derivative portfolio compression[11] in terms of both feasibility and efficiency.

There are numerous studies implementing network methodology on EMIR data. Fiedor (2017) presented a comprehensive description and analysis of the centrally cleared interest rate derivative market in the EU. Bardoscia et al. (2019) build a separate network for each derivative market (interest rate, credit, and foreign exchange derivatives) and, by joining them together in a single multiplex network, they assess the systemic vulnerability of individual institutions and develop centrality measures specifically designed for multiplex networks. Rosati and Vacirca (2020) reconstruct the network of relationships in the centrally cleared derivatives market between clearing members and CCPs, as well as those between clearing members and their clients. They analyse the topology of the network through different centrality measures applied to both the aggregate network and the multiplex network. Then, they identify critical nodes and assess their interconnectedness within and across the different market segments. Drawing on derivative margins data and cash buffers, Bardoscia et al. (2021) develop a model to compute liquidity shortfalls in a stress scenario. Kahros et al. (2021) provide a descriptive analysis of client clearing in both OTC and exchange-traded derivative markets and across all products.

Recently a fourth stream of research appeared, more focused on risk assessment. Among others, Jukonis (2022) introduces a model to estimate market risk for equity derivatives exposures in an adverse scenario.

## 2. Data collection process through trade repositories

In this section, we briefly recall the data collection process. A TR is an entity that centrally collects and maintains records of financial transactions, with the aim of allowing firms to report these transactions to their corresponding regulatory authority. Under Regulations (EU) No 648/2012 (EMIR) and No 2015/2365 (SFTR), TRs collect and maintain records of derivatives transactions. They play a central role in enhancing the transparency of derivative markets and providing detailed data to monitor risks to financial stability.

The EMIR[12] and SFTR[13] regulations require counterparties to report daily to a TR the details of transactions concluded, modified and terminated (life-cycle events) on derivatives and on securities financing transactions, respectively. Once registered, the TR is supervised by ESMA to ensure that it complies on an ongoing basis with all EMIR and SFTR requirements, thereby enabling regulators to

11 Portfolio compression is a multilateral netting technique that enables market participants to coordinate the replacement of existing contracts in order to reduce the size of their mutual obligations, thereby reducing counterparty risk, while maintaining the same underlying market risks.

12 Counterparties subject to reporting on the basis of the EMIR regulation (art. 9) are: financial counterparties (including banks, investment firms, insurance companies, UCITS and pension institutions, central counterparties) and non-financial counterparties (companies established in the EU other than financial counterparties) established in the EU (including their branches and non-EU branches) and all EU branches based outside the EU. Regulation (EU) No 834/2019, which entered into force on 17 June 2019, amended the reporting obligation of derivatives for non-financial counterparties (NFCs): an NFC must report to the TR the OTC derivative contracts belonging to the categories for which it exceeds the thresholds (clearing thresholds) indicated by ESMA, or for all contracts, if it does not calculate its position in relation to these thresholds.

13 Counterparties subject to SFTR reporting (art. 2) are: financial counterparties (including insurance companies, funds and market infrastructures such as central counterparties and central securities depositories) and NFCs established in the EU (including related branches and non-EU branches) and all EU branches of entities based outside the EU.

access data and details of derivative contracts to fulfil their respective mandates. ESMA supervises the TR against which it has authorising, controlling, and sanctioning powers.

The companies can choose the TR to which they report the trades on a daily basis. One company can use more than one TR for EMIR reporting: each counterparty independently reports the concluded transactions to the chosen TR. If other actors are involved (e.g. brokers, CCPs), a single transaction gives rise to multiple reports (see Appendix A.1).

Eight TRs have been registered by ESMA since 2013. After Brexit, only four of them are currently active in the EU market (see Table 1).

TRs registered by ESMA

Table 1

| Denomination | Derivative asset class | Country (currency of the fee) | Effective date |
|---|---|---|---|
| DTCC Derivatives Repository Ltd. (DTI) | all | Ireland (USD) | 14/11/2013 |
| Krajowy Depozyt Papierów Wartosciowych S.A (KDP) | all | Poland (PLN) | 14/11/2013 |
| Regis-TR S.A. (RGS) | all | Luxembourg (EUR) | 14/11/2013 |
| UnaVista TRADEcho B.V. (UVT) | all | Netherlands (EUR) | 14/11/2013 |

Source: ESMA.

It is important that market participants report to TRs all details regarding derivative contracts they have entered into. As a result, information on the risks inherent in derivatives markets will be centrally stored and easily accessible to ESMA, the relevant competent authorities, the ESRB and the relevant central banks of the European System of Central Banks[14]. The provision of TR services is characterised by economies of scale, which may hamper competition in the field. At the same time, the imposition of a comprehensive reporting requirement on market participants may increase the value of the information maintained by TRs and also for third parties providing ancillary services, such as trade confirmation, trade matching, credit event servicing, portfolio reconciliation or compression.

The reports are of two types: (1) trade activity reports, to which we refer to as *activities* (A), containing all reports sent the previous day; (2) trade state reports, to which we refer to as *states* (S), containing all pending trades at the end of the reporting day. While activities show each lifecycle event of a transaction (e.g. conclusion, valuation, modification, termination), states, are at a higher level of aggregation.

Even if data users (e.g. NCAs) can request direct access to TR data, ESMA developed a single access point to EMIR transaction data (i.e. the Trace project) to provide a unified access to the transaction data stored by TRs (see Figure 1). Although the Trace project simplifies data collection by competent authorities, the journey of the reports from the reporting entity to the analyst working with data is far from linear. There is no direct connection between data users and reporting entities and asking for an explanation of a single submission can prove very challenging.

[14] The EMIR and SFTR Regulations and the ESMA regulatory technical standards (EU Reg. 2019/357 and 2019/361) ensure NCAs the visibility of data on the basis of their respective institutional prerogatives. For the Bank of Italy, the access powers are attributable to the following purposes: financial stability, monetary policy, supervision of CCPs, banking and financial intermediaries, and crisis resolution.

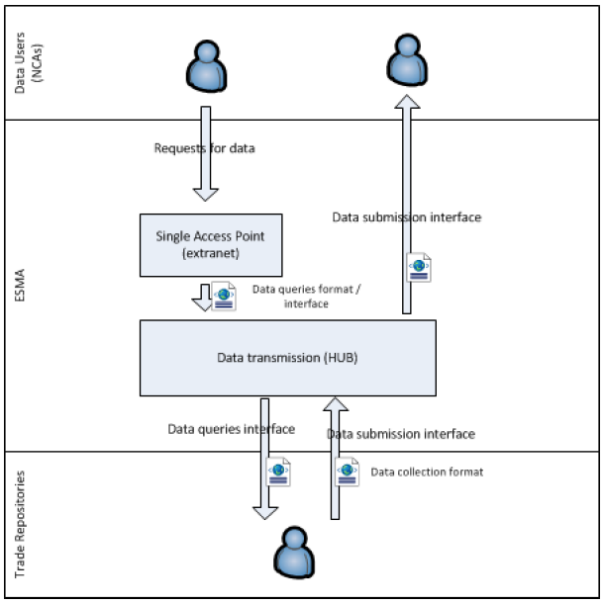



*Figure 1: ESMA Trace project. Source: ESMA documentation.*

# 3. EMIR Data

The focus of this section is on EMIR data and the analyses presented are based on data available to Banca d'Italia as one of the relevant competent authorities. Section 3.1 provides an overview of the main information on the contracts and on the reporting entities. From Section 3.2 to 3.4 we discuss data quality issues, the processing framework and how to extract quantitative information on initial and variation margins. In Section 3.5 we compare, for a sample of Italian banks, the notional amounts extracted from EMIR data to those reported in FINREP and in the supervisory and statistical reports. Market statistics in terms of notional value are reported in Section 3.6. Finally, in Section 3.7 we show examples of networks drawn from EMIR data.

## 3.1 Data overview

Starting from December 29, 2020, Banca d'Italia receives each month through Trace more than half billion daily records collected from four TRs (Figure 2).[15] This means that we have around 30 million records for each trading day; most reports come from Regis. According to Chart 10 of ESMA (2021), Banca d'Italia has access to about half of total EMIR submissions. As reported by ESMA, at the beginning of 2021, the number of EMIR submissions sharply dropped, as UK-based TRs were excluded. Both activities and states are reported for a total of more than 17 billion records in the period from the last days of 2020 to the end of April 2023.

EMIR data are also enriched with other information from Global Legal Entity Identifier Foundation (GLEIF), that allows to identify entities through the legal entity identifier (LEI) code, a 20-digit alphanumeric code that enables clear and unique identification of companies participating in global financial markets.

In more than 70 per cent of the records, both counterparties have an LEI code, that is entities are operating legally as of the last verification. In some cases, however, the LEI code is not available and the reporting entity uses a client code internally defined to identify the counterparties. For instance, this can

[15] In counting the number of reports, we do not consider the updates of the same report but only the last available report.

happen when the counterparty is a retail client (i.e. an individual); there are also cases of firms in which the LEI does not result as *issued* anymore.[16]

In what follows, we provide the broad picture of the EMIR archive stored in the Banca d'Italia big data platform. We focus on the number of submissions (observations) related to *activities* and *states*. Analysing states allows us to have a look at the entire picture of the derivatives reported in EMIR in terms of outstanding contracts. For this reason, Section 3.6 on market statistics deals only with the notional in euro of trade states. In this section instead, we focus on all reports (observations) between January 2021 and April 2023.

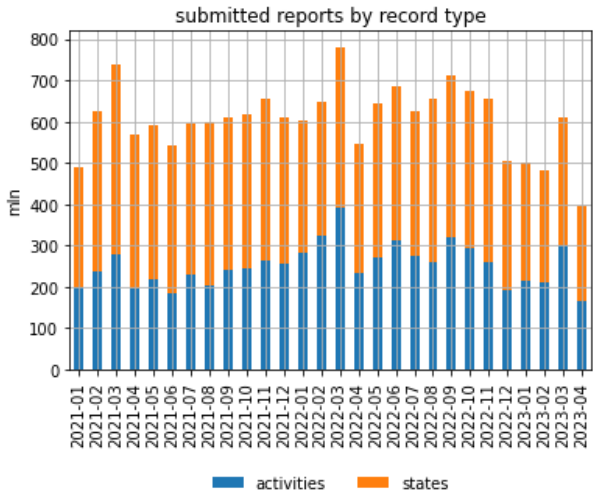



*Figure 2: Overall monthly data. Source: EMIR data available at the Banca d'Italia. Number of monthly reports by record type.*

For both reporting and other[17] counterparty's countries, Figure 3 shows for each country the number of observations (states and activities). Italy is the seventh country by number of reported records. Less than 10 per cent of the transactions involves Italian counterparties. Most of the transactions are reported by Dutch, Cypriot, and German counterparties. Cypriot counterparties report almost 20 per cent of total transactions, but the notional of these trades is small. Even if British counterparties have been excluded from reporting after Brexit, these are still a large portion of the counterparties resulting as "other".

At an individual level, there are thousands of reporting counterparties but the first 50 counterparties represent more than 80 per cent of the observations available at the Banca d'Italia. The counterparties reported as "other" are less concentrated: the first 50 counterparties represent slightly less than 50 per cent of the total observations.

As far as the contract characteristics are concerned (see Figure 4), equity derivatives account for more than 60 per cent of total submissions while contracts with interest rates as underlying are almost 20 per cent. Commodity and currency asset classes are almost equivalent in terms of total observations (around 10 per cent). At least for the counterparties to which we have access to, credit derivatives are a small fraction of the overall derivative market.

16 Possible LEI registration status are: issued, duplicate, lapsed, merged, retired, annulled, canceled, transferred, pending_transfer, pending_archival.

17 On each trade there are two counterparties, i.e. the reporting and the other.

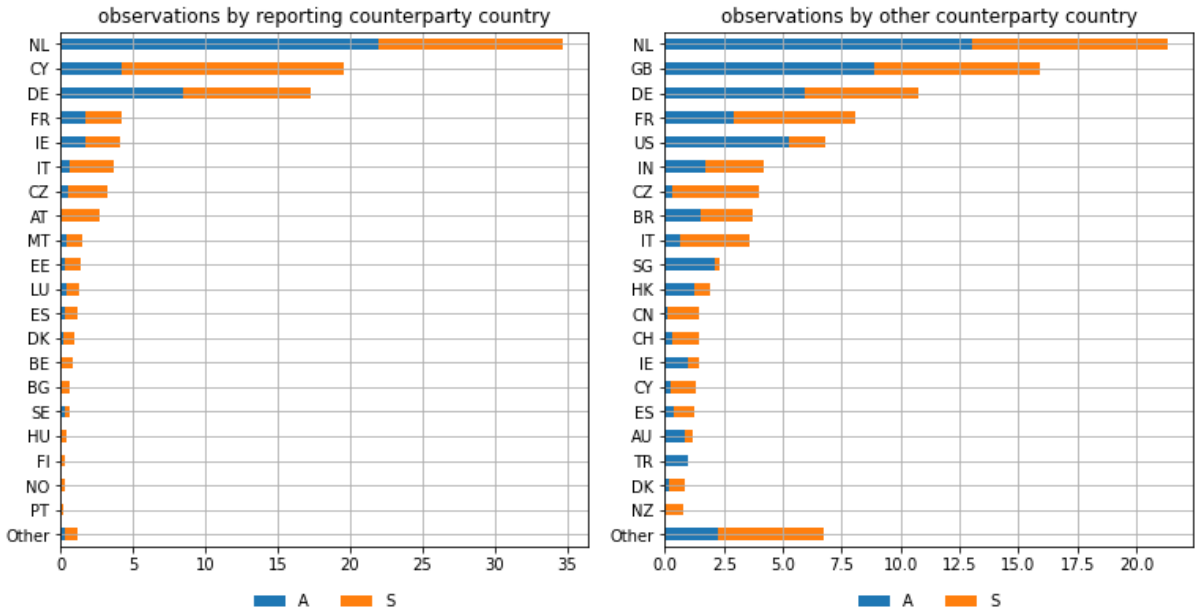



*Figure 3: Overall data by country of the reporting (other) counterparty: number of observations. Source: EMIR data available at the Banca d'Italia. Percentage over the total number of observations.*

More than one-third of reports refer to futures contracts (see Figure 4). Contracts for differences are more than one quarter of the total reports. This is may also be due to the fact that they are retail products and many trades have negligible notional values (see also Figure 7). While options are around 20 per cent of the total, all other contract types, including swaps, are below 6 per cent. Spread bets represent a market niche.

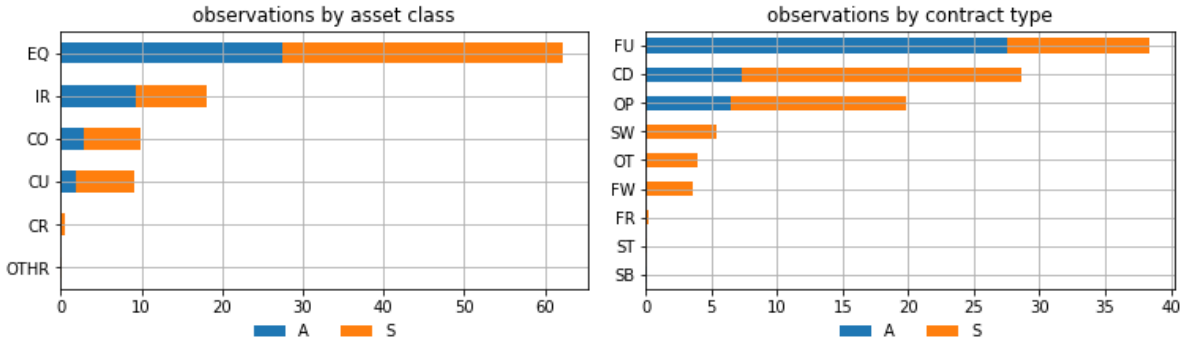



*Figure 4: Overall data by contract characteristics: number of observations. Source: EMIR data available at the Banca d'Italia. Percentage over the total number of observations. Asset classes: commodities (CO), credit (CR), currencies (CU), equities (EQ), interest rates (IR), other (OTHR). Contract types: contracts for differences (CD), forward rate agreements (FR), forwards (FW), futures (FU), options (OP), spread bets (SB), swaps (SW), swaptions (ST), other (OT).*

The main currencies are the euro and the US dollar (see Figure 5). However, certain other currencies are not negligible, at least in terms of number of submissions.[18]

[18] In the case of an interest rate or currency derivative contract, these will be the notional currency of leg 1 and 2.

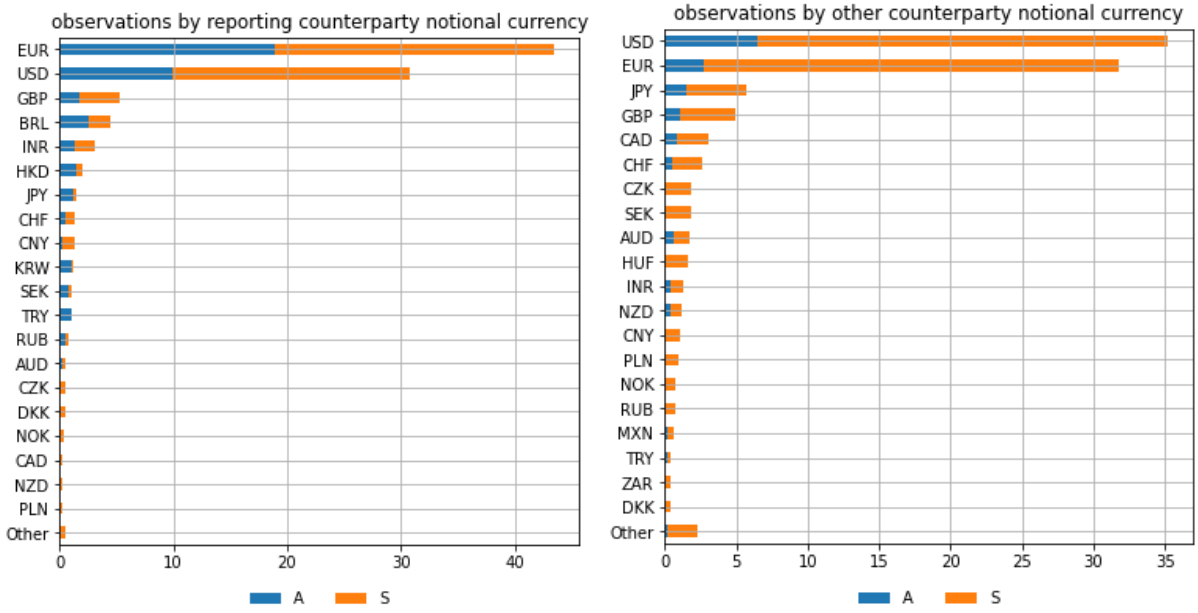



*Figure 5: Overall data by notional currency: number of observations. Source: EMIR data available at the Banca d'Italia. Percentage over the total number of observations.*

While around 63 per cent of overall records on states refer to over-the-counter (OTC) contracts, the remaining 37 are No-OTC, that is they can be exchange traded and cleared (ETD) or traded on multilateral or organised trading facilities (MTF or OTF) and not necessarily centrally cleared.

Figure 6 distinguishes between markets, cleared and non-cleared transactions and collateralisation type. More than one quarter of transactions are uncollateralised. Most of the collateralised transactions are partially or one-way collateralised. It could be the case that the quality of these fields is not always high.

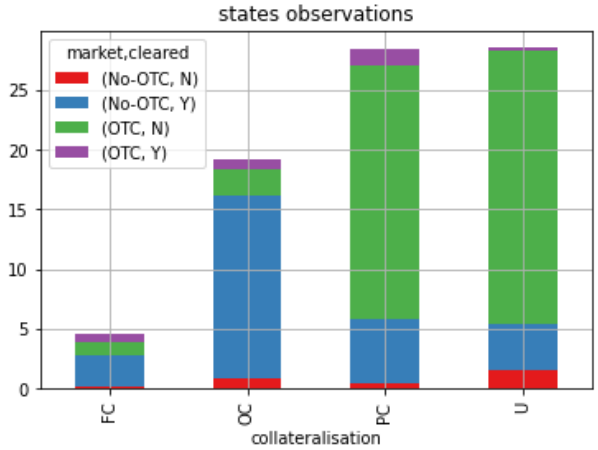



*Figure 6: Overall data by market, clearing and collateral information: trade states observations. Source: EMIR data available at the Banca d'Italia. Percentage over the total number of observations. Collateralisation type: fully collateralised (FC), one way collateralised (OC), partially collateralised (PC), uncollateralised (UC).*

One of the most important quantitative data is the contract notional value. The size of this value depends on various factors, including the counterparties' sectors, the contract type and the asset class. Figure 7 shows the states notional in euro by size buckets. We report "not a number" (*nan*), when the euro exchange rate is not available.[19] Some entities report negative or huge values, which in most cases are reporting errors. More than one quarter of the trades have a notional between 0 and 1,000 euro. Given

[19] The values of this field are not always available for a technical issue we are currently solving.

the volume of trades of this type, it is not always clear whether they are reporting errors or if they are transactions of retail clients on trading platforms. The amount of trades with notional greater than one billion is negligible.

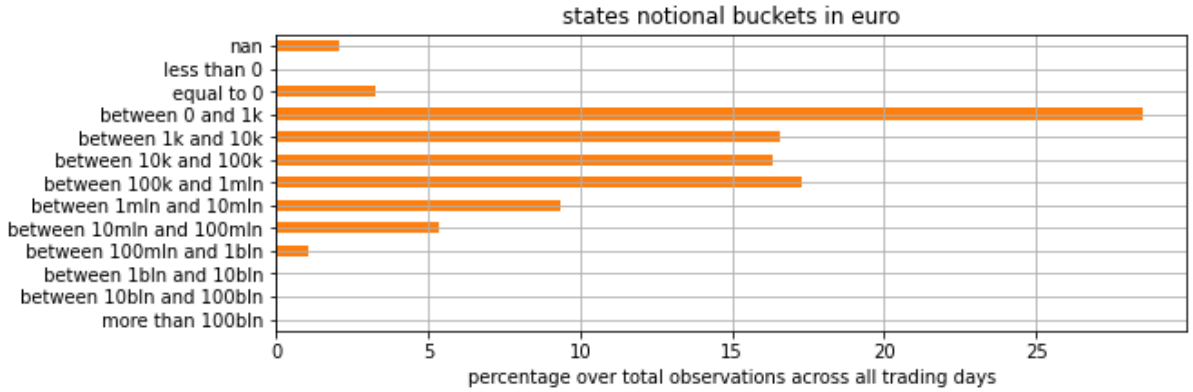



*Figure 7: Trade states notional in euro by size buckets. Source: EMIR data available at the Banca d'Italia. Percentage over the total number of observations.*

## 3.2 Data quality issues

Nearly eight years after the beginning of the reporting mandate, the data quality collected under EMIR is still far from perfect and further effort is needed. The annual reviews of the quality of EMIR data published by ESMA show that, despite the progress made in recent years, certain issues are still present (see ESMA, 2022). In particular, the report published in 2022 shows that, while less than 10 per cent of reported derivatives tend to be reported late by the counterparties, more than 20 per cent do not receive updated valuations on a daily basis, and pairing rate continues to be relatively low at 60 per cent.[20] In a recent letter to the European Commission, the ESRB states that persistent data quality issues pose several limitations to the use of EMIR data for the monitoring of financial stability risks.[21]

Among the issues raised by the ESRB two stand out. First, in the monitoring of the post-Brexit clearing landscape, what appeared to be the repatriation of third-country contracts from the UK to the EU proved to be due to misreported notional values. In addition, many reported transactions lacked the important fields to identify the product class (for example, the underlying rates for interest rate swaps). Second, in monitoring liquidity risks arising from margin calls during the COVID-19 crisis and following the Russian invasion of Ukraine, it appeared that many entities wrongly reported collateral as flows instead of stock values. In addition, the amounts and the portfolio codes do not match those reported by the counterparty.

Similar problems have emerged in the work done at Banca d'Italia. The data quality is not homogeneous across reporting entities, derivatives markets and reported fields. These are the major issues:

1. The classification of the type of players (e.g. dealer vs. investor, hedge fund vs. ETF) is problematic because information on counterparties is limited. On the bright side, LEI codes however allow us to merge EMIR data with external data on the counterparty information that is needed.

[20] See ESMA (2022a).

[21] See ESRB (2022). The letter highlights a few examples of when the EMIR data proved to be fundamental for the ESRB in monitoring developments from a systemic risk perspective but substantial data quality issues hampered the ability to perform the analyses and achieve meaningful and reliable results.

2. The identifier of the derivatives underlying assets may be missing or inconsistently reported (e.g. ISIN vs. differing name(s) in text). Therefore, the aggregation of positions that refer to the same underlying asset is cumbersome.

3. The dual-sided reporting regime reveals in a few cases conflicting information provided by the two counterparties regarding the direction and quantity of the trade and the contract specification. Only 0.3 per cent of total trades in EMIR have a discrepancy in the notional amounts reported by the two counterparties greater than 25 per cent, and in only 0.1 per cent of trades the counterparties provide inconsistent information, by stating that they are both buyers or both sellers at the same time.[22]

4. The usefulness of initial and variation margins for financial stability monitoring is also compromised by the fact that margins are reported at the collateral portfolio level, instead of trade level (see Section 9). The fact that collateral portfolios are specific to a pair of counterparties does not guarantee that the two counterparties include exactly the same trades in their collateral portfolios (e.g. misreporting or data quality issues) and that they use the same name to denote those portfolios. Without the same composition and denomination of collateral portfolios used by the pair of counterparties, the risk of reporting inconsistencies grows.

## 3.3 Data processing framework

In this section, we describe in detail the steps that we take to transition from the raw dataset of "trade state" reports to one that is fit for the analysis. The processing procedure discards erroneous or unnecessary observations (cleaning and filtering), and samples only one report per trade to avoid double counting of a single trade coming from the EMIR dual-sided reported regime (deduplication). We discuss results from the cleaning and deduplication process using "trade state" data sampled at a weekly frequency from December 29, 2020 to August 31, 2022. For each of the weekly data points we select a Friday or the last trading day in the week if it is not Friday. Table 2 outlines the data processing procedure and presents statistics on the median of the weekly results.

The combination of the cleaning and deduplication exercises leaves us with 7.13 million records (42.5 per cent of initial), 7.13 million trades (49.1 per cent of initial), and 86.1 trillion euros of total gross notional (71 per cent of initial). Panel A in Table 2 reports the median daily number of trade state reports (records), number of unique trade identifiers and total gross notional (in euro billions). We call these values *initial*, as they refer to the raw dataset, i.e. before data processing. On an ordinary day, there are almost 17 million records, which refer to approximately 14.5 million distinct trades and have an aggregate notional of 1,261 trillion euro.

In what follows, we assess the impact of the various data processing steps in terms of drop rates computed with respect to the initial values. In order to have a meaningful representation of the effects of the processing procedure on the aggregate notional, we use as initial values the outcome of our first cleaning step, which removes outlier observations characterised by implausible notional amounts (either too small or too large). Net of outliers, the initial total gross notional is just over 121 trillion euros (bottom right corner of Panel A in Table 2).

Panel B in Table 2 illustrates the data cleaning steps and reports the number of records and trade identifiers, as well as the total gross notional that survive and those that are dropped in each step, which are implemented sequentially. Alongside the number of dropped observations, we report its percentage share of the initial value, which we use to gauge the magnitude of the discarded information.

The **first cleaning step** removes erroneous records due to implausible notional values defined as either below 1,000 or above 50 billion euros. These thresholds are based on both standard practices and

[22] The total notional in EMIR affected by each of these inconsistencies is about 1 per cent.

exploratory analysis of the outliers. The choice of the lower bound is in line with previous studies on EMIR data (Abad et al., 2016) and it allows not only to deal with non-positive notional values but also to remove all those trades that given their size might not be important from a systemic risk viewpoint. We note however that these smaller trades account for a substantial part of the raw EMIR dataset. In fact, over 99% of the 5.2 million observations discarded in this first step are records with positive notional value below 1,000 euro and mostly refer to trades that involve online brokerage firms and their retail clientele. While the lower bound is the reason why a third of the initial records and trades are discarded in the first step, the upper threshold is what causes the huge drop in the total gross notional to about a tenth of its initial value. We choose 50 billion, which is bigger than the 10 billion threshold used by Abad et al. (2016), based on evidence that trades with notional amount above that value are never reported by CCPs and are often reported by non-financial counterparties, which makes us confident in treating them as erroneous reporting.

The **second and third cleaning steps** exclude trades that appear inactive. This happens either when the reference date of the reporting is beyond the trade's stated maturity or termination date (Filter 2) or when the valuation of the contract is done more than 100 days from the reference date (Filter 3). The latter filter is more impactful as it drops approximately 7 per cent of both records and trades and 10 per cent of the total gross notional.

The **fourth step** discards reports that do not specify whether the reporting counterparty is buying or selling the derivative. The **fifth step** discards reports where both the reporting entity and its counterparty do not have the LEI code. The impact of these two filters is negligible.

A strong effect on the cleaning procedure comes from the **sixth cleaning step**, which excludes intragroup positions between companies of the same group. We identify these trades using the intragroup flag reported under EMIR, which we further complement with information on group structure and ownership from GLEIF. Intragroup transactions do not constitute standard buy (or sell) market operations; their purpose can rather reflect risk transfer or liquidity management. As such, these positions can be excluded from financial stability analyses. This notwithstanding, they might still be relevant for different kinds of analyses and in this case they might still be looked at. By applying this filter, we drop about 9 per cent of the reports, accounting for almost 8 per cent of the total notional amount in EMIR. It should be noted that this is not a cleaning step but a filter one needs to explore EMIR data from a financial stability perspective, where intragroup positions do not affect the risks of the counterparties.

The **seventh cleaning step** drops redundant records leaving the number of trades unchanged. It ensures that, for a given trade, there exists at most one report per counterparty. This is not data deduplication per se, as we are not choosing which of the two reports is considered, but rather we are focusing on the reports submitted by the same counterparty. The same counterparty can erroneously submit multiple reports for a given trade on a given day for a number of reasons. For instance, this is observed when a report is submitted to flag the origination of a contract and another report is used to provide its valuation.

Based on the list of reported EMIR fields that are recognised as compulsory by the ESMA validation rules, we favour reports with action type[23] of valuation ("V"), i.e. inserted for valuation updates, and drop the others. Whenever the preferred action type "V" is not available, we randomly pick one of the existing reports. [24]

[23] The values of this field can be: new (N), modify (M), error (E), early termination (C), correction (R), compression (Z), valuation update (V) and position component (P). A simpler, from a practical perspective, alternative, giving similar results in terms of total notional values, is to discard all trades having more than one action type for a given reference date and reporting counterparty.

[24] See the Excel file on validation rules at the following link https://www.esma.europa.eu/policy-rules/post-trading/trade-reportingopa.eu

## Cleaning process

Table 2

| | | | | | |
|---|---|---|---|---|---|
| BEFORE CLEANING (A) | | | | | |
| records | initial: | 16,795,372 | | | |
| trades | initial: | 14,525,682 | | | |
| notional (bn) | initial: | 1,261,440 | initial w/o implausible: | | 121,179 |
| AFTER CLEANING (B) | | | | | |
| FILTER 1: implausible notional (<1000 or > 50bn) | | | | | |
| records | survive: | 16,795,372 | drop: | 5,244,802 | (31.2%) |
| trades | survive: | 14,525,682 | drop: | 5,464,352 | (37.6%) |
| notional (bn) | survive: | 1,261,440 | drop: | - | |
| FILTER 2: maturity or termination date before reference date | | | | | |
| records | survive: | 11,472,930 | drop: | 77,641 | (0.5%) |
| trades | survive: | 8,971,797 | drop: | 89,534 | (0.6%) |
| notional (bn) | survive: | 120,884 | drop: | 295 | (0.2%) |
| FILTER 3: valuation date in excess of 100 days from reference date | | | | | |
| records | survive: | 10,341,586 | drop: | 1,131,343 | (6.7%) |
| trades | survive: | 7,956,912 | drop: | 1,014,884 | (7.0%) |
| notional (bn) | survive: | 107,782 | drop: | 13,102 | (10.8%) |
| FILTER 4: missing counterparty side | | | | | |
| records | survive: | 10,341,490 | drop: | 96 | (0.0%) |
| trades | survive: | 7,956,871 | drop: | 42 | (0.0%) |
| notional (bn) | survive: | 107,781 | drop: | 1 | (0.0%) |
| FILTER 5: no LEI for both counterparties | | | | | |
| records | survive: | 10,338,336 | drop: | 3,154 | (0.0%) |
| trades | survive: | 7,952,988 | drop: | 3,884 | (0.0%) |
| notional (bn) | survive: | 107,779 | drop: | 2 | (0.0%) |
| FILTER 6: intra-group transaction | | | | | |
| records | survive: | 8,848,798 | drop: | 1,489,538 | (8.9%) |
| trades | survive: | 7,132,240 | drop: | 820,748 | (5.7%) |
| notional (bn) | survive: | 98,513 | drop: | 9,266 | (7.6%) |
| FILTER 7: multiple reports of the same trade submitted by the same counterparty | | | | | |
| records | survive: | 8,842,711 | drop: | 6,088 | (0.0%) |
| trades | survive: | 7,132,240 | drop: | 0 | (0.0%) |
| notional (bn) | survive: | 98,403 | drop: | 109 | (0.1%) |
| FILTER 8: unmatched pair of counterparties in double reporting | | | | | |
| records | survive: | 8,833,348 | drop: | 9,364 | (0.1%) |
| trades | survive: | 7,130,328 | drop: | 1,912 | (0.0%) |
| notional (bn) | survive: | 98,290 | drop: | 113 | (0.1%) |
| AFTER CLEANING and DEDUPLICATION (C) | | | | | |
| records | final: | 7,130,329 | (42.5%) | | |
| trades | final: | 7,130,329 | (49.1%) | | |
| notional (bn) | final: | 86,093 | (71.0%) | | |

The **final cleaning step** excludes trades for which reports submitted by distinct counterparties are inconsistent in the identification of the pair of entities involved. This can happen for instance when both A and B are entities subject to the EMIR reporting obligation but while A reports a trade with B, B reports that the same trade is in place with a third entity C. We treat these cases as erroneous observations to drop. The impact of this filter is very limited.

Panel C in Table 2 shows the results of deduplication of the clean EMIR dataset, by dropping one observation per pair of matched transactions. Deduplication of two-sided reported trades is carried out according to a pecking-order criterion, centred around reporting reliability. More specifically, we assume that CCPs authorised and recognised by ESMA are more reliable than their clearing members, and the latter are more reliable than other entities. After obtaining the list of authorised[25] and recognised[26] CCPs from the ESMA's website, we retrieve their clearing members by sampling the non-CCP counterparties of their trades. We distinguish 3 cases of trades that need deduplication: (i) for trades between a CCP and its clearing member, we take the CCP report; (ii) for trades between a clearing member of one of the CCPs and a counterparty that is neither a CCP nor a clearing member of it, we take the clearing member's report; (iii) for all other trades (i.e. none of the counterparties is a CCP's clearing member or they both are) we take the report of the counterparty that on the date of interest has submitted the largest number of reports.

## 3.4 Dealing with collateral portfolios

To minimise the costs of collateralisation, the management of the collateral related to derivatives contracts between two counterparties is usually performed by grouping contracts with the same characteristics or underlying (i.e. collateral portfolios). By following this market practice, initial and variation margins are reported at the collateral portfolio level, rather than at the contract level. However, while at the level of a single contract the counterparties submit data by reporting the same trade identifier (*trade id*), this is not the case for the collateral portfolio code. Indeed, there is not a unique identifier, and two counterparties may label the same collateral portfolio with different codes. Data show that two pairs of counterparties may not agree on the number of collateral portfolios that exist between them. It can occur that, for a given trade, the first counterparty reports a collateral portfolio code, while the second counterparty does not. It can also occur that only one counterparty submits the report of a given trade.

To deal with these issues, we leverage the double reporting of the same trade and define, on each reference date, a unique collateral portfolio identifier composed by the LEIs of the counterparties sorted in alphabetic order and a progressive number that denotes and groups together all collateral portfolios that are used by the two counterparties and have some trades in common. On the basis of this identifier, for each counterparty pair it is possible to compute the initial and variation margin posted and received by a given entity at a given reference date.

In order to extract all available information provided by the double reporting, the analysis of margin should be conducted considering both sides of the reporting. For each counterparty pair and each unique collateral portfolio, it is possible to merge, by means of a selected function, the values of margins received by a counterparty with those posted by the other counterparty.[27] In practice this means that an analyst working with EMIR data needs both a deduplicated data set to investigate notional and contract values and a not deduplicated data set to examine margins.

Margin values are not always available in EMIR data. Firstly, according to Regulation (EU) No 148/2013, most non-financial corporations are not required to report collateral of contracts. Secondly, the EMIR

25 https://www.esma.europa.eu/sites/default/files/library/ccps_authorised_under_emir.pdf

26 https://www.esma.europa.eu/sites/default/files/library/third-country_ccps_recognised_under_emir.pdf

27 Average, median or maximum between the two reported values can be selected.

validation rules by ESMA do not always allow retrieving the values of margins posted by an entity from the trade reports submitted by its counterparty. For instance, in the case of partially collateralised derivative contracts, where the reporting counterparty might be the only one receiving the initial margin, the validation rules do not require populating the field of initial margin received. Without this value, it is not possible to know the value of the initial margin posted by the other counterparty unless this submits its own report. For all these reasons, any analysis of margins should hinge on EMIR considering both sides of the reporting and give priority to the reports submitted by the entity whose margins are being assessed.

## 3.5 Comparison with other sources

For Italian banks EMIR is not the only available source of data on exposure to derivatives (but it is so for non-financial firms). Italian banks have to submit information regarding their exposures to derivatives in FINREP and in the national supervisory and statistical reports (Matrice dei conti).[28] We consider these two datasets to validate the notional amounts extracted from EMIR data available at Banca d'Italia and obtained after the cleaning and deduplication phases.

Supervisory and statistical reports have a quarterly frequency and include information on notional values and carrying amounts of derivatives used for hedging and trading purposes. The values are broken down by type of underlying instrument.[29] FINREP framework requires banks to quantify their exposures both at group and at individual level, while the national supervisory reports collect individual data only.

The cross-check has been carried out on six banks considered at a group level and on a quarterly basis.[30] Each bank has a total notional exposure of more than 100 billion. These banks account for almost the entire derivatives exposure held by the Italian banking system. Figure 8 reports, for each bank, the comparison between the three data sources in terms of notional amounts held at the end of each quarter between Q4 2020 and Q4 2022.

[28] On 1 January 2014, the fourth directive on capital requirements set a regulation of reporting to harmonise the collection of data in the Eurozone. The FINREP report is among those designed to meet the fourth directive legislative requirements.

[29] The derivatives are distinguished among interest rates derivatives, equity derivatives, currency derivatives, credit derivatives and commodity derivatives. There is also a residual class for other types of derivatives.

[30] Given that the national supervisory report is at the individual level, the notional values have been aggregated according to the bank group structure at the relevant date.

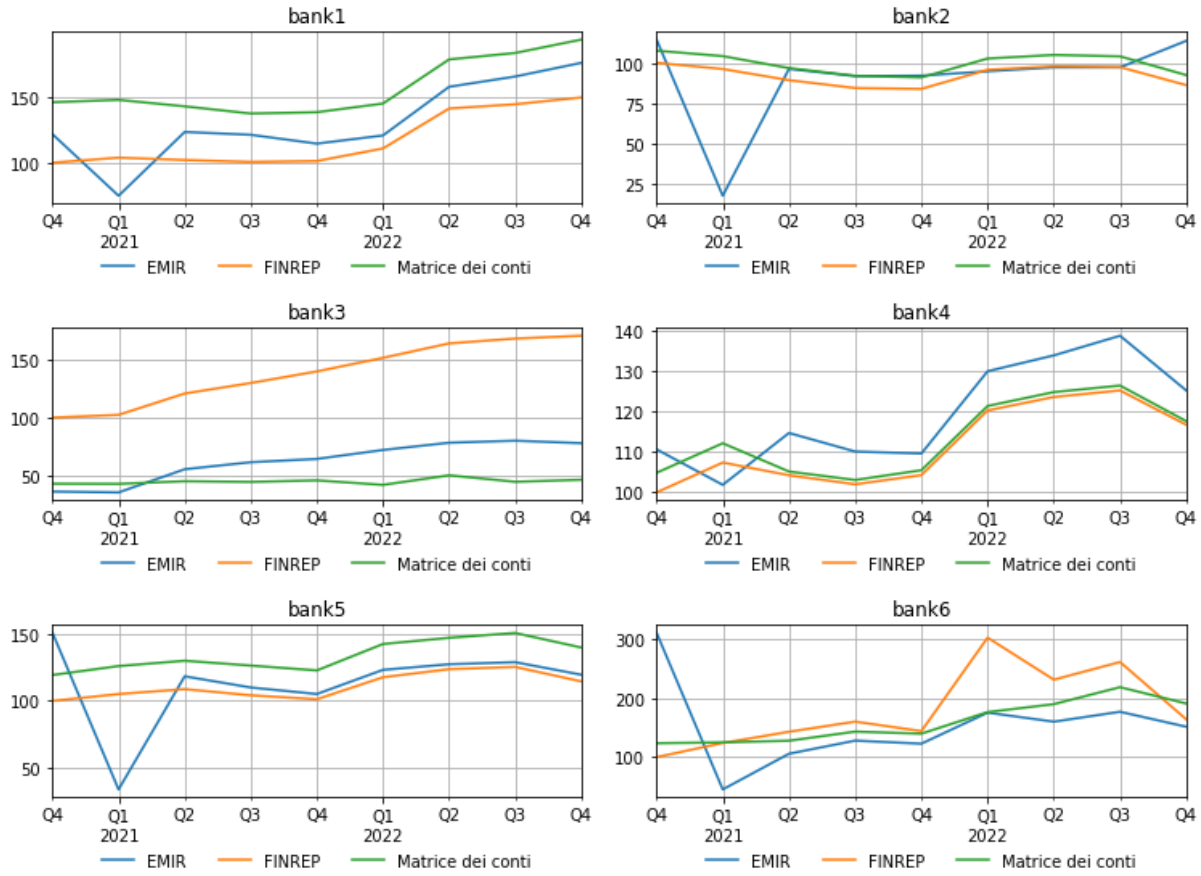



*Figure 8: Comparison between different sources between Q4 2020 and Q4 2022. Source: deduplicated EMIR data available at the Banca d'Italia, FINREP and supervisory and statistical reports (Matrice dei conti). All data are rescaled so that, for each bank, the FINREP value is equal to 100 at the Q4 2020. For a bank we consider the reports submitted by the banks including the transactions with the parent company.*

Except for bank3, the amounts reported in EMIR are close to those reported in the other two datasets. Moreover, dynamics are very similar. Values reported by bank3 in FINREP are much different from those contained in EMIR and in the national supervisory report. This is due to problems in properly identifying the derivatives held by subsidiaries of the group of this bank.

Table 3 provides an overview of the notional amount (in billion euro) of derivatives, by asset class, at the end of 2022. The values are rescaled for confidentiality reasons: for each bank, the FINREP value is equal to 100. Interest rate derivatives are the most relevant. Without considering bank3, the reported total notional values of the three different sources are similar. The differences by asset class could be due to different classification criteria between EMIR and the other two sources.

Overall, it seems that the total notional amounts of derivatives held by the main Italian banks obtained by using different sources of information are reasonably close to each other. The main differences are therefore related to banks characterised by a high degree of complexity.

EMIR, FINREP and supervisory and statistical reports data as of December 31, 2022

Table 3

| | CO | CR | CU | EQ | IR | OTHR | TOT |
|---|---|---|---|---|---|---|---|
| bank1 | - | 0.0 | 3.0 | - | 114.6 | - | 117.6 |
| | - | 0.0 | 4.8 | 0.0 | 95.2 | - | 100.0 |
| | - | 0.0 | 4.7 | - | 124.6 | - | 129.4 |
| bank2 | 0.4 | 0.1 | 1.0 | 6.0 | 124.3 | - | 131.8 |
| | 0.3 | 1.4 | 1.3 | 4.6 | 92.4 | - | 100.0 |
| | 0.7 | 1.5 | 2.7 | 3.4 | 98.9 | - | 107.2 |
| bank3 | 1.0 | 0.2 | 1.3 | 1.4 | 41.9 | 0.0 | 45.9 |
| | 0.5 | 0.4 | 6.1 | 1.7 | 91.2 | 0.2 | 100.0 |
| | 0.1 | 0.0 | 1.9 | 0.2 | 25.2 | - | 27.4 |
| bank4 | 0.2 | 4.3 | 9.2 | 1.9 | 91.6 | - | 107.3 |
| | 0.2 | 4.3 | 5.5 | 1.1 | 88.9 | - | 100.0 |
| | 0.2 | 4.3 | 5.3 | 1.1 | 89.9 | 0.0 | 100.8 |
| bank5 | 0.7 | 0.0 | 24.8 | 7.3 | 71.4 | - | 104.4 |
| | 0.1 | 0.0 | 22.2 | 6.2 | 71.4 | - | 100.0 |
| | 0.1 | 0.0 | 22.6 | 9.1 | 90.3 | - | 122.2 |
| bank6 | 0.0 | 17.5 | 3.2 | 8.4 | 63.2 | - | 92.3 |
| | 0.6 | 17.2 | 5.8 | 8.4 | 68.1 | - | 100.0 |
| | 2.1 | 29.9 | 5.2 | 9.2 | 69.9 | 0.0 | 116.3 |

Source: EMIR (first row), FINREP (second row) and supervisory and statistical reports (Matrice dei conti, third row) data available at the Banca d'Italia. The total notional of derivatives contracts is reported. All data are rescaled so that, for each bank, the FINREP value is equal to 100 at the Q4 2022. Asset classes: commodities (CO), credit (CR), currencies (CU), equities (EQ), interest rates (IR), other (OTHR).

## 3.6 Market statistics

Following the data processing framework defined in Section 3.3 we compute market statistics based on state reports.

Figure 9 shows the notional of outstanding derivatives contracts by asset class and contract type. We evaluate the monthly averages of daily information on the cleaned and deduplicated data. Figures are lower than those reported in ESMA (2021), as Banca d'Italia has access to only a portion of EMIR data. However, Banca d'Italia can also access information not involving Italian counterparties (see Section 3.1).

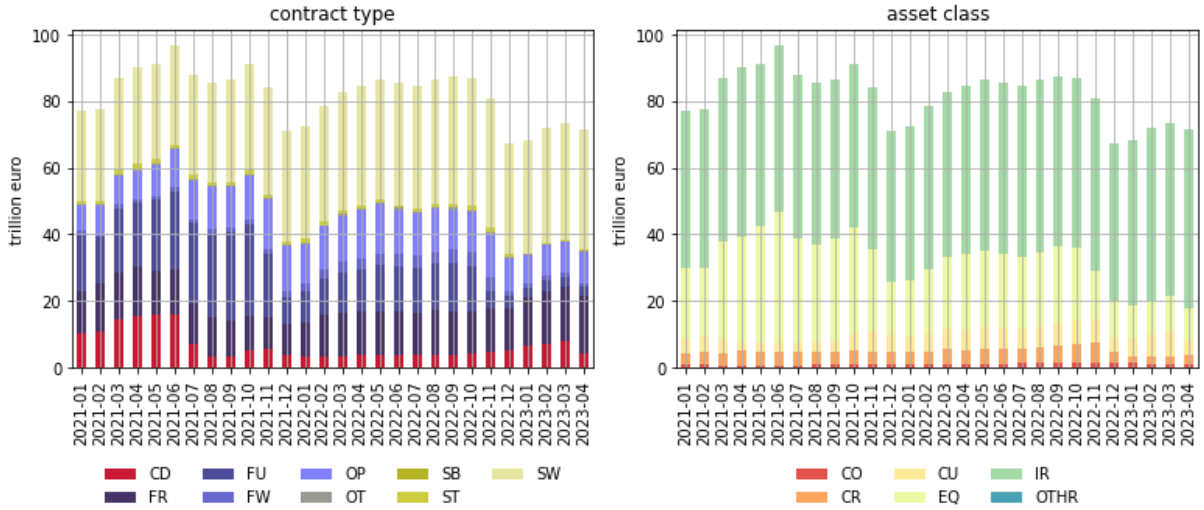



*Figure 9: Trade states notional in euro by asset class and contract type. Source: EMIR deduplicated data available at the Banca d'Italia. Monthly averages of daily data. Asset classes: commodities (CO), credit (CR), currencies (CU), equities (EQ), interest rates (IR), other (OTHR). Contract types: contracts for differences (CD), forward rate agreements (FR), forwards (FW), futures (FU), options (OP), spread bets (SB), swaps (SW), swaptions (ST), other (OT).*

The market size seems quite stable over the observation period, even if a decrease has been registered in the last few months. Regarding the contract types, excluding contracts for differences, their size does not vary significantly. As for other financial assets, the overall notional of the derivative portfolio of major financial intermediaries, which are the main EMIR contributors, varies slowly.

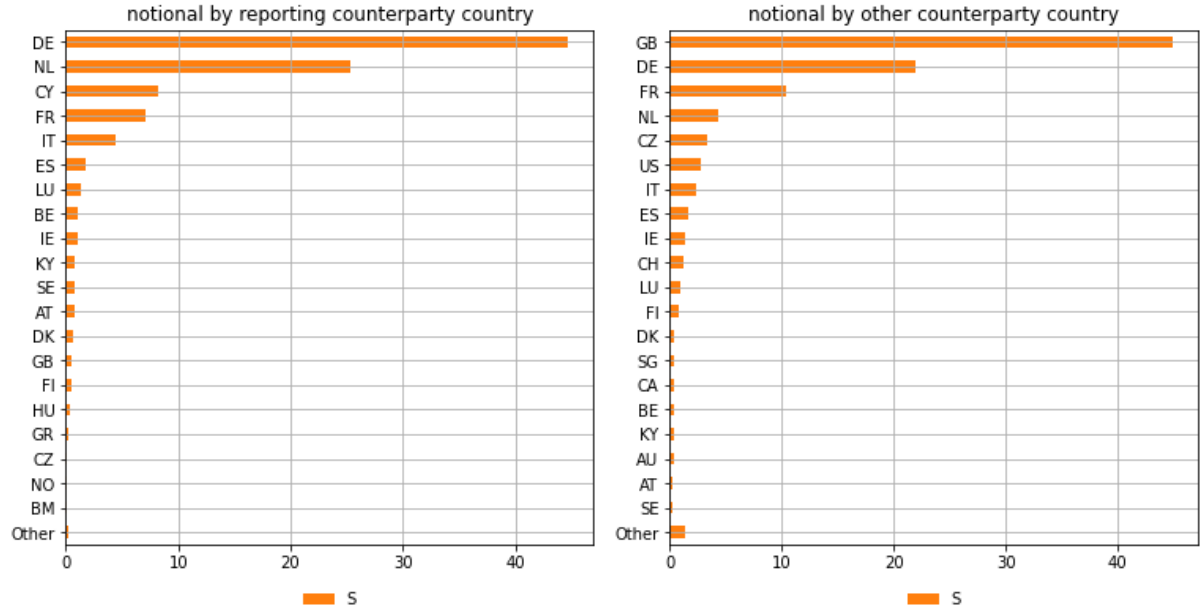



*Figure 10: Overall data by country of the reporting (other) counterparty: states notional. Source: EMIR deduplicated data available at the Banca d'Italia. Percentage over the total states notional.*

In terms of notional (see Figure 10), German entities are the major players in the market, both as reporting and as other counterparty. Italian entities are below five (three) per cent of the total states notional as reporting (other) counterparty. This may be due to the fact that major CCPs on derivatives are not Italian. The reporting of cleared derivatives involves multiple entities, including the CCP, and for this reason these entities are among the major counterparties both in terms of reports and of notional (see Appendix A.1). According to the data available at the Banca d'Italia, in Europe, the derivative market is concentrated in a few countries and British counterparties are among the major players.

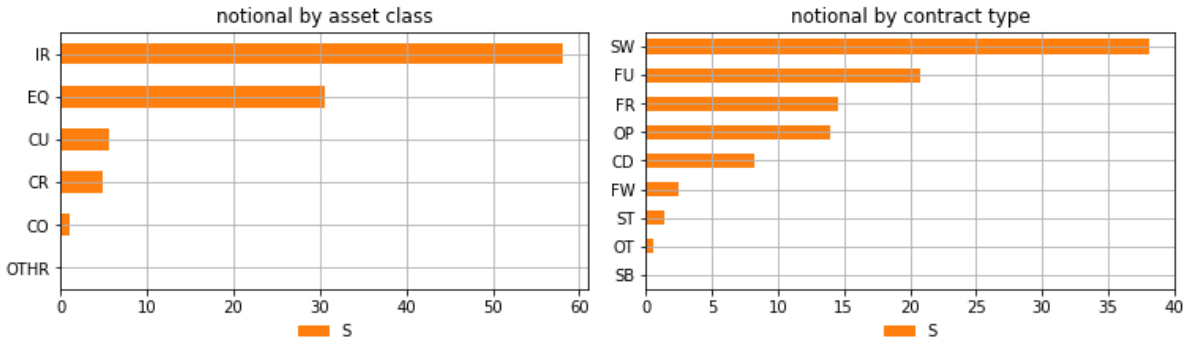



*Figure 11: Overall data by contract characteristics: states notional. Source: EMIR deduplicated data available at the Banca d'Italia. Percentage over the total states notional. Asset classes: commodities (CO), credit (CR), currencies (CU), equities (EQ), interest rates (IR), other (OTHR). Contract types: contracts for differences (CD), forward rate agreements (FR), forwards (FW), futures (FU), options (OP), spread bets (SB), swaps (SW), swaptions (ST), other (OT).*

Although equity derivatives account for more than 60 per cent of total submissions, these represent less than one-third of the overall notional (see Figure 11). More than half of the total notional refers to contracts on interest rates, mainly swaps. Credit and currency derivatives are almost equivalent in terms of overall notional, but not in terms of observations (see Figure 4). At least for the counterparties to which we have access to, commodity derivatives are a small fraction of the total notional on derivatives. Future contracts represent one-fifth of the market in terms of notional. Options and FRAs are slightly below 15 per cent. Contracts for differences are more than one-quarter of the total reports but account for less than ten per cent of the total notional. The notional of swaptions is small. This is also the case for other derivatives. Spread bets are negligible.

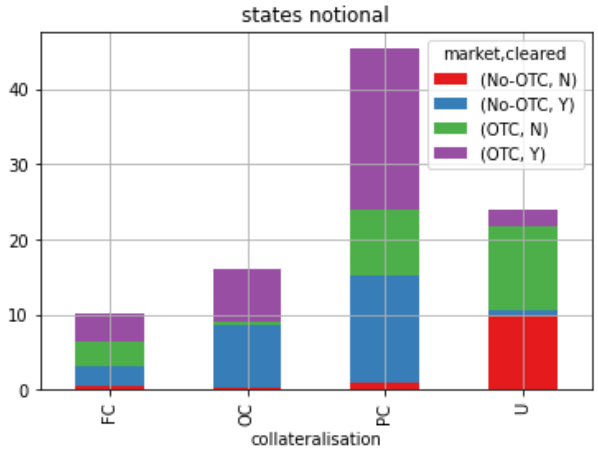



*Figure 12: Overall data by market, clearing and collateral information: states notional. Source: EMIR deduplicated data available at the Banca d'Italia. Percentage over the total states notional. Collateralisation type: fully collateralised (FC), one way collateralised (OC), partially collateralised (PC), uncollateralised (UC).*

In terms of notional, around 61 per cent of trade states refer to OTC contracts, and the remaining 39 to are not OTC. However, three fourth of the contracts have some form of collateralisation and two third are cleared (see Figure 12).

## 3.7 Some EMIR network snapshots

Another way to exploit EMIR data is by building network graphs to visualise interconnections between market participants. Figure 13 shows a network based on a centrality measure. First, we select only the

connections (edges) between counterparties greater than 500 million, then we choose only the entities (nodes) with more than 250 billion euro derivative exposures at the end of September 2021 (2022).

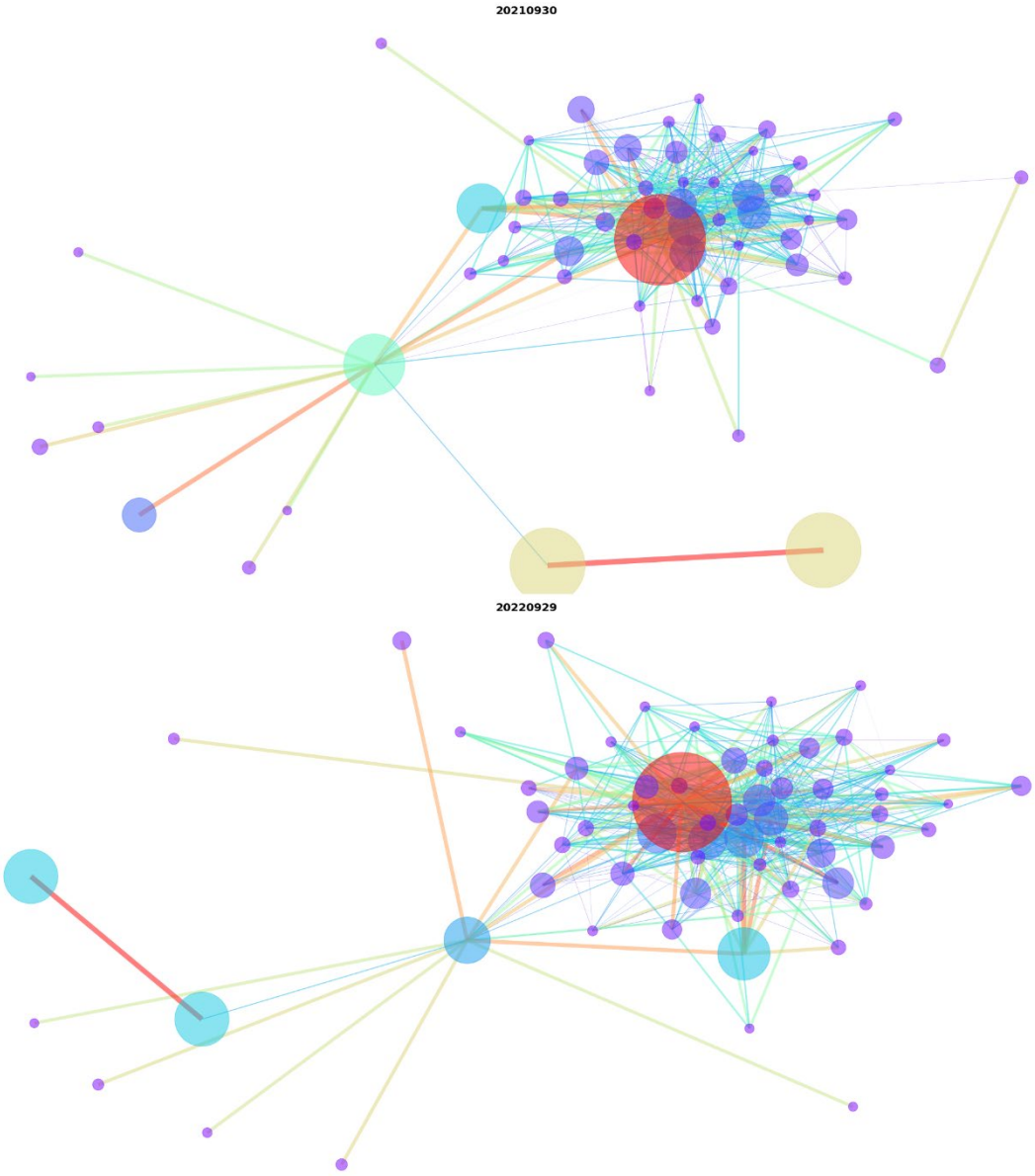



*Figure 13: network of the biggest connections among major reporting counterparties at September 30, 2021 and September 29, 2022. Source: EMIR deduplicated data available at the Banca d'Italia.*

The size and the colour of the edge represent how large the notional of the transactions is among entities, while the size and the colour of the circle represent the total notional of derivative transactions of the entity. We are considering different contract types and there is more than one CCP in the network. A CCP is the entity with the largest notional (the red circle) and is placed in the centre of the graph. The network varies slowly over time.

## 4. Conclusion

The complexity of EMIR is twofold: derivatives are complex instruments and there is a huge amount of data available to be organised and analysed. EMIR has required central banks, supervisors and regulators to adopt innovative data analysis approaches: "big data" are to be collected and processed with state-of-the-art technologies. This represents a challenge, but also an opportunity.

After briefly reviewing the main features of the EMIR regulation and the research leveraging this data source, we describe the journey of the data from the reporting entities to the data users and we illustrate the EMIR data available at the Banca d'Italia, with a look towards financial stability analysis and research. We provide a data overview and present a data processing framework to deal with data quality issues on the notional value and to define the identifier needed to compute initial and variation margins from data on collateral portfolios. Similar approaches could be considered to solve possible issues regarding the contract value. Finally, we provide some market statistics and an example of how a simple EMIR network could be visualised.

## References


Abad, J., Aldasoro, I., Aymanns, C., D'Errico, M., Rousová, L. F., Hoffmann, P., Langfield, M., Neychev, M., Roukny, T. (2016). Shedding light on dark markets: First insights from the new EU-wide OTC derivatives dataset. ESRB Occasional Paper Series, 11.

Abbassi P., and Bräuning, F. (2021). Demand effects in the FX forward market: Micro evidence from banks' dollar hedging. *The Review of Financial Studies*, 34(9), 4177-4215.

Abraham, L. (2020). Competition analysis on the over-the-counter credit default swap market. *arXiv preprint arXiv:2012.01883*.

Ascolese, M., Molino, A., Skrzypczynski, G., Cerniauskas, J. and Pérez-Duarte, S. (2017), Euro-area derivatives markets: structure, dynamics and challenges, *IFC-National Bank of Belgium Workshop on* "Data needs and Statistics compilation for macroprudential analysis"

Bianchi, B. (2021). Cross-border credit derivatives linkages. ESRB: Working Paper Series 2021/115.

Bardoscia, M., Bianconi, G., and Ferrara, G. (2019). Multiplex network analysis of the UK over-the-counter derivatives market. *International Journal of Finance & Economics*, 24(4), 1520-1544.

Bardoscia, M., Ferrara, G., Vause, N., and Yoganayagam, M. (2021). Simulating liquidity stress in the derivatives market. *Journal of Economic Dynamics and Control*, *133*, 104215.

Bellia, M., Girardi, G., Panzica, R., Pelizzon, L., and Peltonen, T. A. (2019). The demand for central clearing: to clear or not to clear, that is the question. SAFE Working Paper, 193.

Bias, D., Guagliano, C., Haferkorn, M., Haimanna, M., and Kaserera, C.. Mutual funds and derivatives: Evidence from linked fund-trade data. *Preprint*.

Bouveret, A. and Haferkorn, M. (2022). Leverage and derivatives – the case of Archegos. ESMA Report on Trends, Risks and Vulnerabilities Risk Analysis.

Cielinska, O., Joseph, A., Shreyas, U., Tanner, J., and Vasios, M. (2017). Gauging market dynamics using trade repository data: the case of the Swiss franc de-pegging. Bank of England Financial Stability Paper, 41.

Cominetta, M., Grill, M., and Jukonis, A. (2019). Investigating initial margin procyclicality and corrective tools using EMIR data. ECB *Macroprudential Bulletin*, *9*.

Dalla Fontana, S., Holz auf der Heide, M., Pelizzon, L., and Scheicher, M. (2019). The anatomy of the euro area interest rate swap market. ECB Working Paper Series, 2242.

D'Errico, M., and Roukny, T. (2021). Compressing over-the-counter markets. *Operations Research*, 69(6), 1660-1679.

ESMA (2021). EU Derivatives Markets, ESMA Annual Statistical Report 2021.

ESMA (2022a). EMIR and SFTR data quality report 2021.

ESMA (2022b). Guidelines for reporting under EMIR.

ESMA (2023). 2022 report on quality and use of transaction data.

ESRB (2020). Mitigating the procyclicality of margins and haircuts in derivatives markets and securities financing transactions. ESRB Expert Group on the Macroprudential Use of Margins and Haircuts.

ESRB (2022). ESRB's view regarding data quality issues and risks for financial stability.

Fiedor, P., and Killeen, N. (2021). Securitisation special purpose entities, bank sponsors and derivatives. *Journal of International Financial Markets, Institutions and Money*, 75, 101452.

Fiedor, P., Lapschies, S., and Országhová, L. (2017). Networks of counterparties in the centrally cleared EU-wide interest rate derivatives market. ESRB Working Paper Series, 54.

Grothe, M., Pancost, N. A., and Tompaidis, S. (2021). Empirical analysis of collateral at central counterparties. ESRB Working Paper Series, 131.

Kahros, A., Pioli, A., Carraro, T., Gravanis, M., & Vacirca, F. (2021). A Descriptive Analysis of the Client Clearing Network in the European Derivatives Landscape. *Journal of Financial Market Infrastructures*, 9(1).

Kenny, O., Killeen, N., and Moloney, K. (2016). Network analysis using EMIR credit default swap data: micro-level evidence from Irish-domiciled special purpose vehicles (SPVs). *IFC Bulletin*, 41.

Joseph, A., & Vasios, M. (2022). OTC Microstructure in a period of stress: A Multi-layered network approach. *Journal of Banking & Finance*, 138, 106400.

Jukonis. A. (2022). Evaluating market risk from leveraged derivative exposures. ECB Working Paper Series, No. 2722.

Lenoci, F. D., & Letizia, E. (2021). Classifying Counterparty Sector in EMIR Data. In *Data Science for Economics and Finance* (pp. 117-143). Springer, Cham.

Levels, A., van Stralen, R., Kroon, S., and van Lelyveld, I. (2018). CDS market structure and risk flows: the Dutch case. DNB Working Paper, n. 592.

Mazzacurati, J., Guagliano, C., and Spolaore, A. (2021). Funds and single-name CDS: Hedging or Trading, ESMA Working Paper.

Pedro, S. A. B. (2020). An assessment of the portuguese OTC market network structure under EMIR-Bringing light to the portuguese OTC CDS Market. Master thesis.

Pérez-Duarte, S., and Skrzypczynski, G. (2019). Two is company, three's a crowd: automated pairing and matching of two-sided reporting in EMIR derivatives' data. Ninth IFC Conference on "Are post-crisis statistical initiatives completed?".

Rosati, S., & Vacirca, F. (2020). Interdependencies in the euro area derivatives clearing network: A multi-layer network approach. *Journal of Network Theory in Finance* 5(2), 1–41.

Schroeder, F., Lepone, A., Leung, H., and Satchell, S. (2020). Flash crash in an OTC market: trading behaviour of agents in times of market stress. *The European Journal of Finance*, 26(15), 1569-1589.

Ullersma, C., and van Lelyveld, I. (2022). Granular data offer new opportunities for stress testing. *Handbook of Financial Stress Testing*, ed. Farmer, J.D., Kleinnijenhuis, A.M., Schuermann, T., Wetzer, T., pp.185-207.

van Lelyveld, I. (2017), The use of derivatives trade repository data: possibilities and challenges, *IFC-National Bank of Belgium Workshop on* "Data needs and Statistics compilation for macroprudential analysis"

# Appendix

## A.1 Reporting of centrally cleared trades

Under EMIR, the reporting and the number of entities involved in a trade depend on whether the trade is uncleared (Figure A.1) or centrally cleared (Figure A.2). These are illustrative examples.

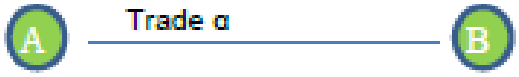



*Figure A.1: uncleared derivatives. Source: Eurex documentation "Reporting by Eurex Clearing according to EMIR Article 9".*

If a trade α between A and B is uncleared, there is a direct contractual relationship between the two counterparties (Figure A.1). If both counterparties have an obligation under EMIR, the bilateral trade must be reported by both sides, A (B) as the reporting counterparty of the trade and B (A) as the other counterparty. This is the standard case for over-the-counter (OTC) derivatives trades that are not cleared by a central counterparty (CCP).

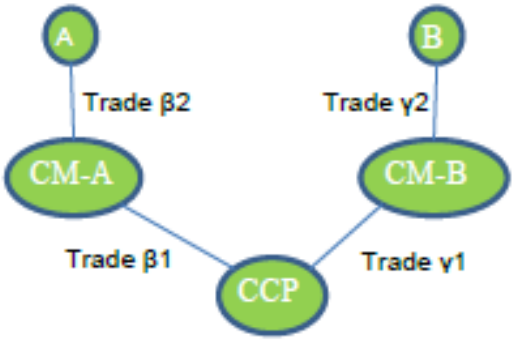



*Figure A.2: exchange traded and OTC cleared derivatives. Source: Eurex documentation "Reporting by Eurex Clearing according to EMIR Article 9".*

When instead a trade is centrally cleared, the reporting involves multiple entities. In the example of Figure A.2, A and B – both clients – are not clearing members of the CCP. Each of them refers to a clearing member, namely CM-A and CM-B. The latter can trade with the CCP. If the derivative is traded on an exchange (e.g. a futures contract), the orders of A and B are matched on the trading venue immediately, resulting in cleared trades. However, there is neither a direct link between A and B nor a direct contractual relationship between A, B and the CCP. In fact, in order for A (B) to have a centrally cleared trade, CM-A (CM-B) concludes a trade β1 (γ1) with the CCP, which therefore is the counterparty of the trade. CM-A (CM-B) also concludes a back-to-back trade β2 (γ2) with A (B). If all entities are subject to reporting obligations under EMIR, all trades (β1, β2, γ1 and γ2) are reported. All these trades share a field that identifies the presence and the name of the CCP.[31] The same reporting framework as shown in Figure A.2 also applies to OTC cleared derivatives, with the only difference being that there is no trading venue that matches the orders of A and B.

Finally, in the special case of clearing members that want their trades cleared by a CCP, the reporting is similar to that in Figure A.2, which means that the CCP appears as the counterparty of the clearing members in their trades.

[31] In EMIR reporting there exists a field (i.e. reporting tracking number) that links all trades (β1, β2, γ1 and γ2) together. Unfortunately, most of the time the information is missing and therefore a link between the trades cannot be established.

## A.2 Data processing and confidentiality issues

At the Banca d'Italia the management of the EMIR database is carried out by the Statistical Data Collection and Processing Directorate (RES) with the help of the IT function. The EMIR database is stored in a Hadoop cluster and can be queried using Spark, a unified analytics engine for large-scale data processing. The data processing can be executed using the Jupyter notebook and implemented in Python (PySpark) or R (SparkR). As an alternative, users can implement their code in SQL. The results of the processing phase can be investigated on the same notebook or can be exported to be analysed through the most common econometric or statistical software.

The users should be aware that EMIR information must be treated as confidential according to art. 83 of the EMIR. Furthermore, this archive is subject to Regulation (EU) No 2016/679 concerning the protection of individuals in relation to the processing of personal data as the possibility that this archive may contain personal data is not excluded, although this is not required by law. Therefore, those who want to access EMIR information must comply with the constraints imposed by the regulations and provisions relating to the protection of personal data.

In particular:

- the users are subject to official secrecy pursuant to art. 7 of the Consolidated Law on Banking and Credit (Legislative Decree no. 385 of 1 September 1993);
- access can only take place for the pursuit of the objectives to perform an institutional function;
- no confidential information can be provided to unauthorised third parties or to external subjects or bodies;
- the data used in the context of analyses, research works and publications can only be disclosed in aggregate form, in order to avoid the possibility that the parties involved in the operations are identified.

## A.3 EMIR introspection for the adjustment of GLEIF data

Entities in EMIR are represented by their LEI code. To identify these entities and their group, EMIR data are enriched with the information provided by the Global Legal Entity Identifier Foundation (GLEIF). GLEIF information includes, among other things, the legal name of the entities, the legal address and the group structure (direct and ultimate parent company), if any. The key of the enrichment is the LEI code.

Since we noticed some errors in GLEIF data, particularly on the group structure, we tried to adjust them by leveraging the EMIR intragroup field. In this field, each reporting entity can select 'Y' for intragroup transaction and 'N' for non-intragroup transactions. This means in practice that, given a pair of counterparties, if 'Y' is reported both the reporting counterparty and other counterparty should belong to the same group[32]. In theory, is it possible to identify the group structure (i.e. all LEI codes of entities belonging to the same group) by looking directly on intragroup trades and without considering GLEIF data. We refer to this approach as *EMIR introspection*. However, since usually not all entities in a group are involved in derivative trading, to obtain a more reliable group structure, we consider both the information coming from EMIR and those coming from GLEIF. Finally, for all intermediaries supervised by the Banca d'Italia, we consider the information on the group structure as collected for supervisory purposes.

[32] We noticed some errors in the reporting of this field and for this reason, if an entity results with more than one parent company, we decided to select as parent only the company having with the given entity the largest number of intragroup transaction ('Y'). All possible ultimate parent companies are preliminary selected on GLEIF. Additionally, we exclude all investment funds from the group structure, even if a 'Y' is reported. The list of investment funds is retrieved from the ECB web-site.

# EMIR data for financial stability analysis and research

Dario Ruzzi
(joint work with M.L. Bianchi and B. Sorvillo)
Banca d'Italia
Financial Stability Directorate

**3rd IFC and Bank of Italy Workshop on**
*Data Science in Central Banking: Enhancing the access to and sharing of data*
October 18, 2023

# Outline

- EMIR data
- access to data
- use of data
- data cleaning (including *deduplication*)
- data quality on notional values
- data quality on the value of contract
- research works on EMIR data

# EMIR data in a snapshot



# EMIR data

The European Market Infrastructure Regulation (EMIR), adopted in 2012, requires all EU counterparties to report their **derivative transactions** to trade repositories (TRs).

The information collected in EMIR can be categorized as follows:

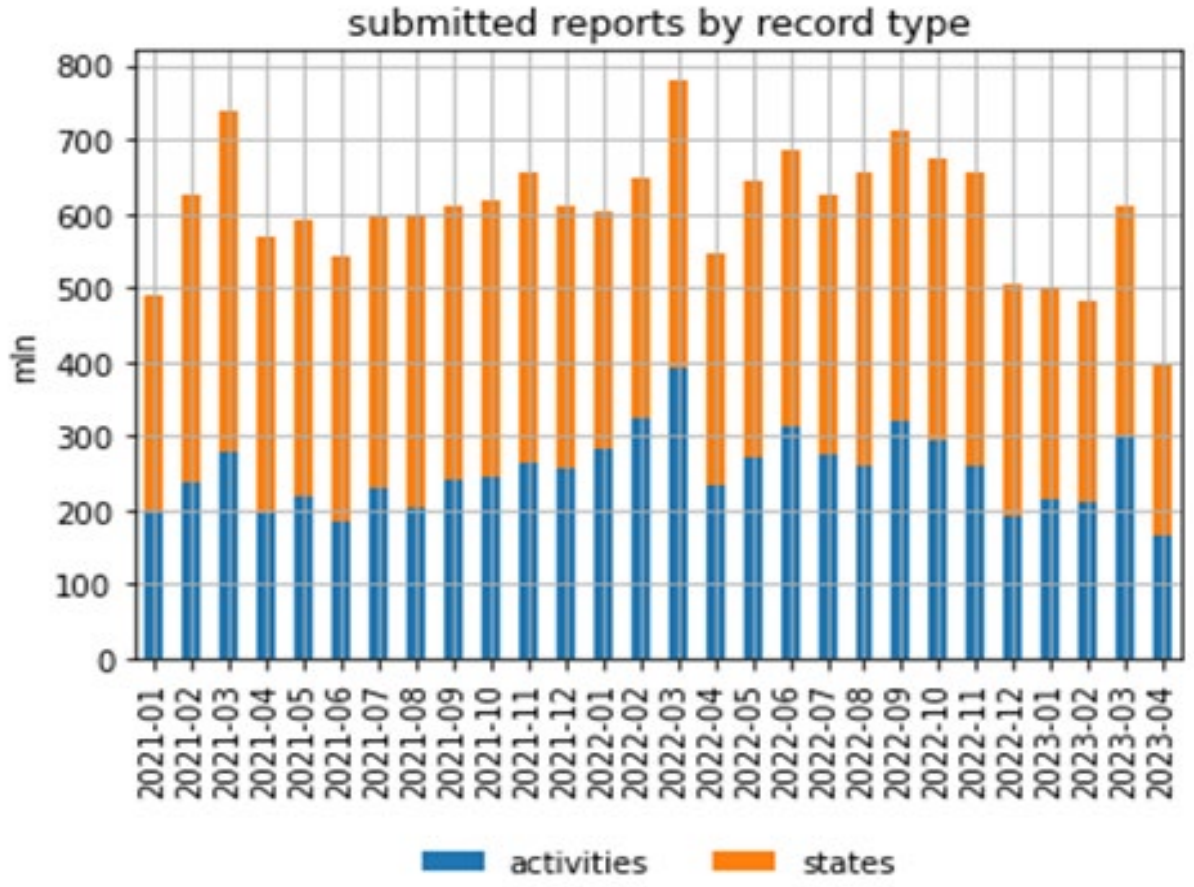
- *trade details*: detailed information related to derivative transactions, including trade identifiers, dates and times of trade execution, product type, underlying assets, notional amounts, contract maturity dates, contract characteristics;
- *counterparty information*: counterparties involved in derivative transactions, including the legal entity identifier (LEI) and their roles in the transaction;
- *valuation and margin data*: valuation of derivative contracts and of the related margins on a daily basis

# EMIR data

The reports are of two types:

1) trade activity reports or ***activities*** (*flows*) containing all reports sent the previous day;
2) trade state reports or ***states*** (*outstanding positions)*, containing all pending trades at the end of the reporting day.

For risk analyses purposes “states” are usually considered.



# Access to data






- Starting from 29 december 2020, we collect EMIR transaction data from ESMA TRACE (4 trade repositories).
- Around 600 million observations each month (states + activities)
- Data available at *t*+2 (*t*+3)
- EMIR users work with «enriched» data and/or other «big» dataframes (Gleif, Banca d'Italia internal databases on securities or entities, ESMA FIRDS)
- cluster Hadoop/Spark (go live: July 2021) : PySpark (*), SparkR, SAS or SQL
- PySpark via Jupyter «is the way»

# Use of data

- Users in various Directorates, mostly in *Financial Stability*
- Deep dives and internal notes on specific counterparties, underlyings or products
- Banca d'Italia **product intervention power** (MiFIR and MiFID II) to assess the financial stability risks associated with financial instruments, structured deposits and related financial activities/practices
- Global and cross country systemic risk monitoring (FSB and NCAs)
- Various research projects (more details in the following slides)
- … market risk sensitivity analysis

# Some data quality issues

- **Reporting errors:** terminated transactions, unresonable notional or contract value, wrong sector and intragroup fields
- **Data gaps:** some TRs do not send us all data as per the regulation
- **Inconsistencies**: side and valuation provided by counterparties of the same trade
- **Fair value:** not always representing the position value (e.g. futures)
- **Initial and variation margins (IM and VM)**:
  - reported at *collateral portfolio* level (do we see all trade in a collateral portfolio?)
  - reporting requirements for posted vs received depend on collateralization type
- **Transaction chain**: it is extremely difficult to reconstruct the entire transaction chain of a contract (centrally cleared trades): the report tracking number does not help much

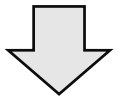
# Data cleaning: our steps

1. remove implausible notional < €1000  and > €50 bln (-38% trades)
2. exclude inactive trades (-7.6%)
3. exclude trades where buyer and seller are not clear (≈ 0%)
4. remove trades where both counterparties have no LEI (≈ 0%)
5. exclude intragroup positions based on Gleif, Banca d'Italia supervisory registers, and ECB list of investment funds (-5.7%)
6. keep only one report per date-counterparty-trade (≈ 0%)

# Deduplication: our approach

Deduplication of two-sided reported trades is carried out according to a pecking-order criterion, centered around reliability and number of reports:

1) CCPs (more reliable than their clearing members, CMs)
2) CMs (more reliable than their clients)
3) All other counterparties



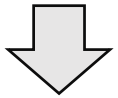
cleaned deduplicated data

ready for the analysis

# Dealing with collateral portfolios

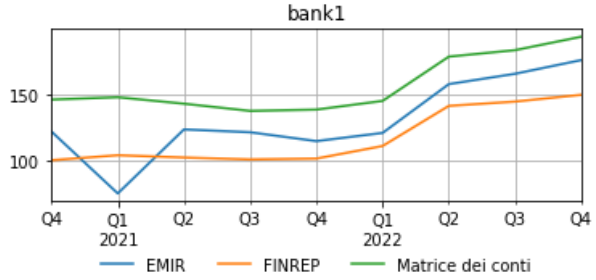
Margins are reported at the collateral portfolio level:

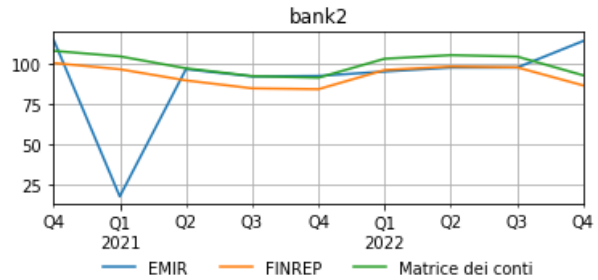
1) no unique portfolio identifier shared by both counterparties
2) the content of portfolios may differ between counterparties
3) some collateralization types (e.g. partial) do not require reporting of received margins

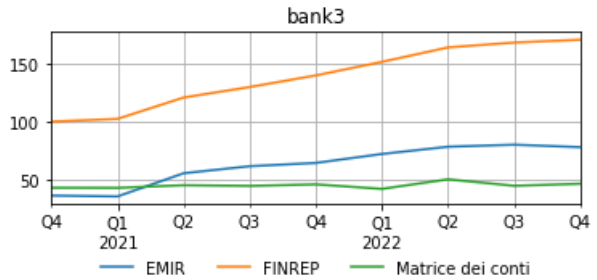


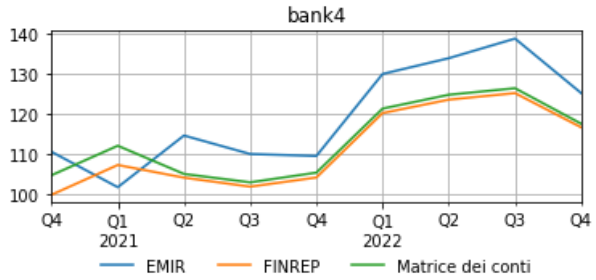
we NEED the double reporting here!

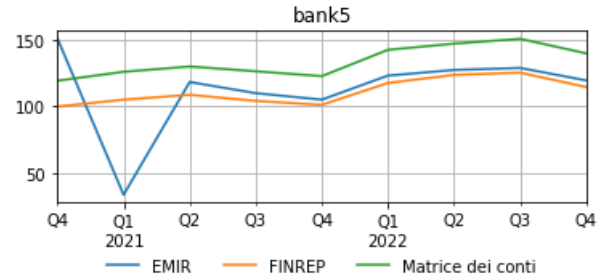
# Data quality on notional values: EMIR vs Finrep

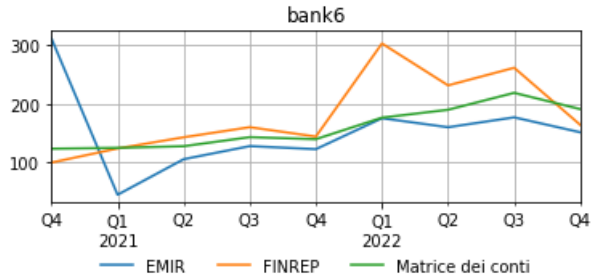


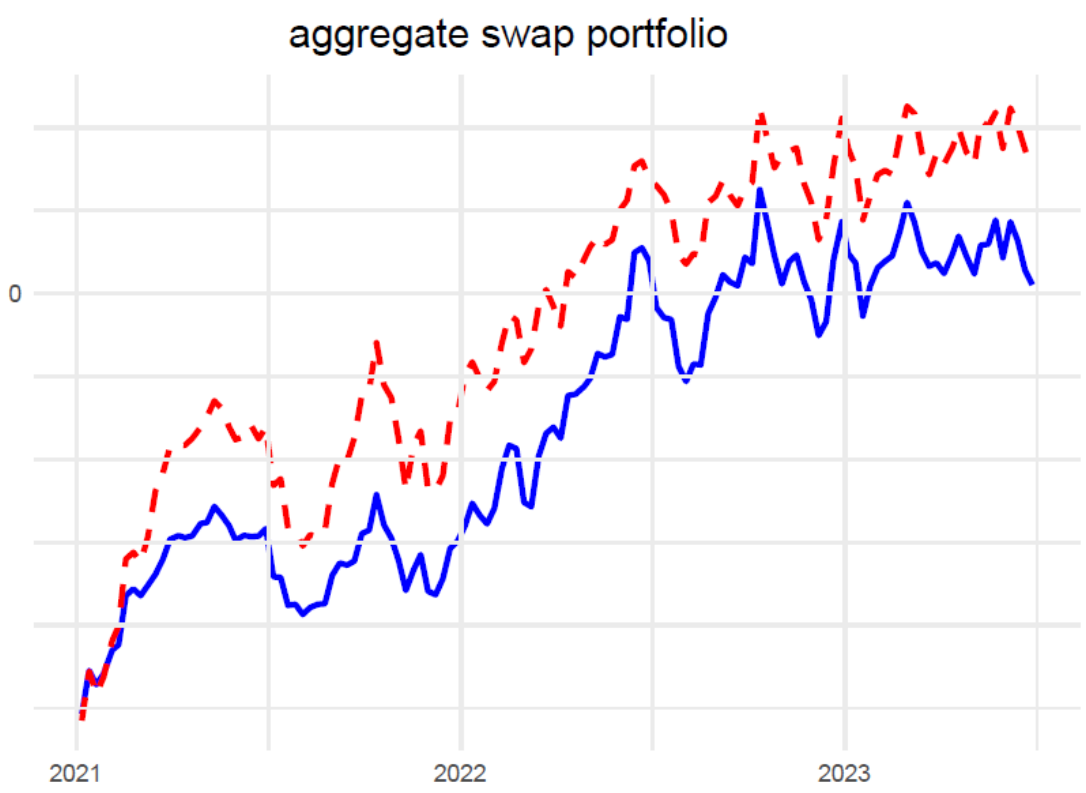



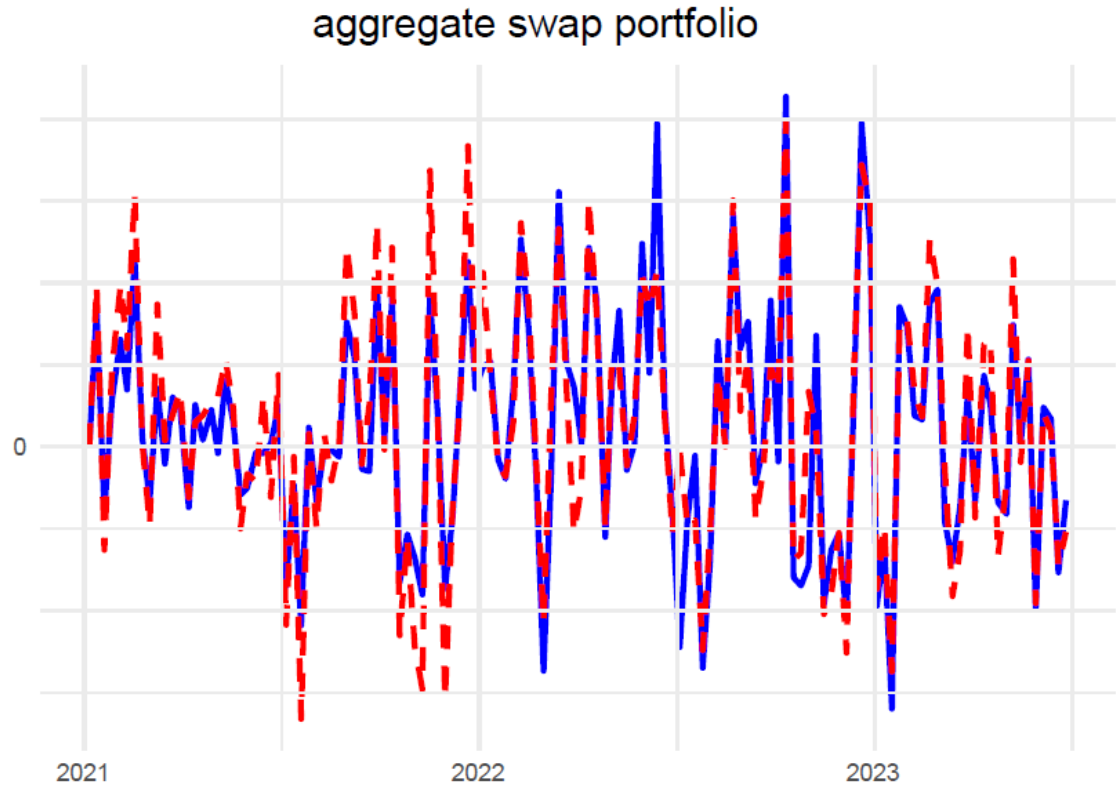




- Source: cleaned deduplicated EMIR data available at the Banca d'Italia, FINREP and supervisory and statistical reports (Matrice dei conti). All data are rescaled so that, for each bank, the FINREP value is equal to 100 at the Q4 2020.

# Data quality on the value of contract: IR swaps

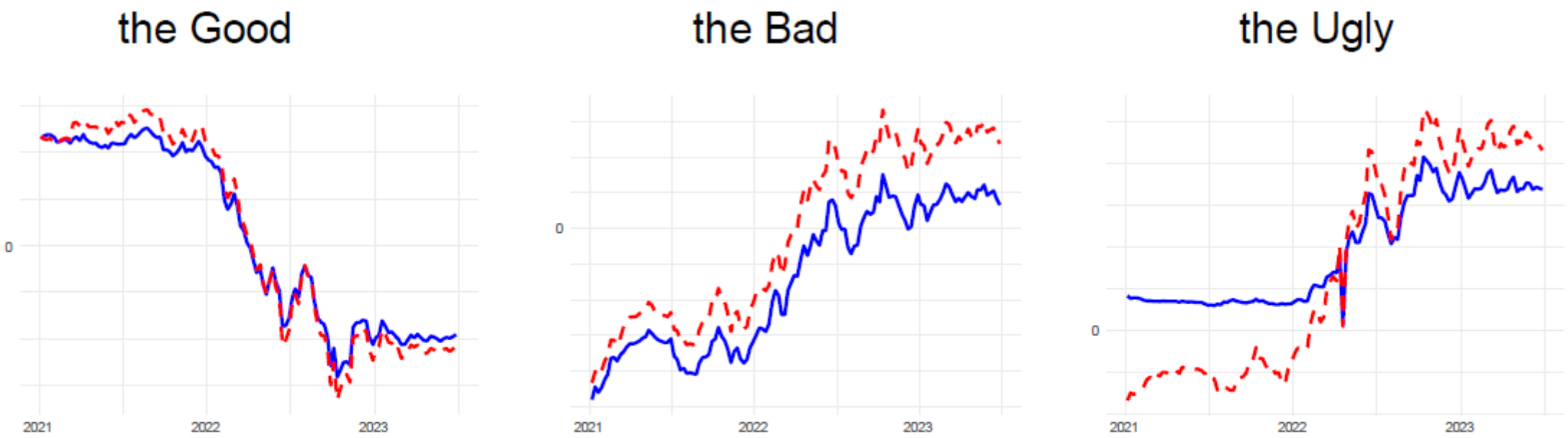



swap portfolio observed in EMIR and implied by the model




swap portfolio observed in EMIR and implied by the model

# Data quality on the value of contract: IR swaps




swap portfolio observed in EMIR and implied by the model

# Research works on EMIR data at Banca d'Italia

- *Shifting the yield curve for Italian banks*: valuation of interest rate derivative contracts and risk sensitivity analysis using granular information on these derivatives and fixed-rate bonds (*available upon request*)
- *$CO_2$ emission derivatives In Italy*: analysis of the market of $CO_2$ emission permits derivatives in Italy (*available upon request*)
- *EMIR networks*: investigation of the structure of EMIR networks (notional and margins), their dynamics over time and the main risk drivers (*work in progress*)
- *Interest rate futures*: analysis on sovereign bond futures (*work in progress*)
- ... a lot of work

Thank you for your attention