Using trade-level derivatives data for macroprudential analysis

IPS 3, Evolving statistics in support of central bank policies

Authors: Gemma Agostoni, Annalaura Ianiro, Francesca Lenoci, Grzegorz Skrzypczynski (European Central Bank, Directorate General Statistics / Directorate General Macroprudential Policy & Financial Stability)

Executive summary

G20 leaders pledged at the Pittsburgh summit in September 2009 to reform OTC derivatives markets to improve their transparency, prevent market abuse and reduce systemic risks. The European response to this commitment is the European Market Infrastructure Regulation (EMIR), which entered into force in 2012. EMIR imposes several requirements on entities that enter derivative contracts, such as the implementation of risk management standards, clearing of certain classes of derivatives through central counterparties (CCPs) and extensive reporting obligation.

In line with EMIR, daily granular information on derivative contracts has to be reported to dedicated trade repositories (TRs), which are then obliged to share subsets of this information with more than 100 authorities in the EU, including the European Central Bank (ECB). Since the beginning of the reporting obligation in 2014, EMIR data have proven instrumental in carrying out various ECB tasks.

However, users still face challenges that prevent them from fully exploiting the dataset, due to the mixed data quality of the information reported, the complexity of underlying contracts, and the quickly evolving derivative product landscape. While the robustness of EMIR reporting has improved significantly since its inception in 2014, thanks to the combined efforts of regulators, TRs and market participants, reporting errors are still a significant barrier for analysts working with these data.

Data science techniques are necessary to detect and address data quality issues in large-scale datasets, such as EMIR. In this spirit, the ECB has designed and implemented a solution to automatically detect, log, and interpret the developments in granular datasets in the presence of potential anomalies, with a direct application to EMIR. On top of that, the ECB developed a set of in-depth cleaning procedures, specific for EMIR data, aimed at producing a ready-to-use dataset for systemic risk identification and financial stability assessments.

Leveraging on the resulting cleaned dataset, the ECB analyses and monitors the euro area derivatives markets, identifies risks to financial stability and draws relevant policy conclusions. For instance, during the 2020 Covid-19 market turmoil, information on margins reported in EMIR has proved fundamental to understand the liquidity stress derivatives' users were facing and the spill-over to other entities[1]. More recently, EMIR data was used to closely follow the developments in the energy derivatives markets[2] and gauge the risks that prolonged high volatility in these markets could pose to financial stability.

This paper describes the data science tools and techniques developed at the ECB to make possible and facilitate the use of granular EMIR information for systemic risk assessments. It also presents some key examples of how EMIR data has been used in the broader context of financial stability analysis, using the abovementioned solutions. Despite the improvements achieved, the quality of the reporting is still not optimal, and the examples demonstrate the need for further work on the application of data science tools to granular largescale datasets of financial data. The application of such techniques would lead to a more efficient and widespread use of these precious sources of information among regulators.

[1] Published in the ECB's November 2021 Financial Stability Review, see Carraro, et al. (2021)[2] Published in the ECB's November 2022 Financial Stability Review, see Furtuna, et al. (2022)

[2] Published in the ECB's November 2022 Financial Stability Review, see Furtuna, et al. (2022)

Table of Contents

1. Introduction

The financial crisis of 2007 – 2008 clearly showed that the complexity and opacity of the derivative markets can lead to materialization of systemic risk. At the Pittsburgh summit in September 2009 the G20 leaders pledged to reform the OTC (over-the-counter)[3] derivative markets to "… improve transparency, mitigate system risk, and protect against market abuse". The measures to be implemented included trading of standardised contracts on exchanges or electronic trading platforms, as well as their clearing by central counterparties (CCPs), higher capital requirements for non-centrally cleared contracts, and finally reporting of derivatives to trade repositories (TRs).[4]

Since then, 19 out of 20 FSB[5] member jurisdictions implemented comprehensive trade reporting requirements.[6] The European response to the above commitment was the European Market Infrastructure Regulation (EMIR)[7], which entered into force on 4 July 2012. The data collected under the regulation is accessible to various EU authorities, supporting their tasks and mandates. At the ECB, the EMIR data has been intensively used for tasks related to financial stability monitoring, market infrastructures oversight, micro-prudential supervision, among others.

The paper is organized as follows. Section 1 briefly describes the EMIR regulation and its reporting framework and lays out the main challenges associated with using the data. Section 2 outlines the processes at the ECB aiming at regular monitoring of data collection and processing, data quality assurance, and data cleaning. Section 3 presents some example applications of the EMIR data to macroprudential analysis conducted by the ECB. Section 4 concludes and presents authors' views on the future of the dataset.

1.1. What is EMIR?

The EMIR regulation imposes several requirements on EU entities entering derivative contracts, including clearing of certain classes of derivatives through CCPs, introduction of risk-mitigation techniques for bilateral OTC derivatives, and obligation to report the details of the derivative contracts and their lifecycle. The counterparties are obliged to report granular information on transaction-by-transaction level. The scope of the data to be reported includes both counterparties-specific information as well as the details of the contract, for example information on various entities involved in the trade, product and underlying identification, notional amounts, daily contract valuation and margins, as well as a range of asset class specific variables.[8]

EMIR mandates the EU counterparties[9] to report the information described above to TRs, specialized entities registered and supervised by European Securities and Markets Authority (ESMA). The TRs are then obliged to share relevant subsets of this information with more than 100 authorities in the EU, depending on their individual mandates.[10] This includes the ECB, which is entitled to receive data on the trades executed by euro area entities or written on euro area underlying instruments.

---

[3] I.e. not traded on an exchange, agreed bilaterally between the two parties.
[4] See: G20 Leaders Statement: The Pittsburgh Summit, http://www.g20.utoronto.ca/2009/2009communique0925.html
[5] Financial Stability Board
[6] See FSB (2022)
[7] Regulation (EU) No 648/2012 of the European Parliament and of the Council of 4 July 2012 on OTC derivatives, central counterparties and trade repositories, see: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02012R0648-20210213
[8] EMIR technical standards specify 129 fields for reporting, see also here: https://www.esma.europa.eu/data-reporting/emir-reporting
[9] Private individuals, as well as certain intragroup transactions concluded by non-financial companies, are excluded from the reporting obligation.
[10] The scope of data accessible to specific authorities is defined in the Regulation (EU) No 151/2013.

While the global G20 guidance refers to OTC derivatives, EMIR goes a step further and mandates also the reporting of exchange-traded contracts – differently to the Dodd-Frank Act in the US. Another feature distinguishing EMIR from its US counterpart is the double-sided reporting obligation, i.e. the requirement for both counterparties to the derivative to report its details to the TRs.

1.2. EMIR main challenges

EMIR reporting obligation entered into force on 12 February 2014, and since the beginning it posed considerable challenges for the authorities using the data to fulfil their mandates.[11] Although the quality of reporting has improved over the years, in several circumstances EMIR users might have faced the doubt of whether reported data were outliers or referring to real transactions. In 2014, EMIR was a completely new reporting framework for complex, heterogenous and evolving product landscape, and the reporting guidance encompassed in the EMIR technical standards, was not sufficient to guarantee a uniform reporting, leaving many areas open for interpretation.[12] Furthermore, the inconsistent reporting occurring overtime by the same entity might suggest that some counterparties did not extensively develop neither a robust and efficient IT infrastructure nor a comprehensive processes to validate their reporting and quality assurance[13]. In this way, the abovementioned limitations have led to multiple data quality issues, making the data hardly usable. Furthermore, the regulation did not impose on the TRs any specific formats regarding the structure of reporting to the authorities, making the aggregation of the information reported by TRs prone to inconsistencies and time-consuming.

Since then, many successful initiatives to improve the quality of the EMIR data have been undertaken. In 2014 and 2015, ESMA mandated the TRs to carry out a pre-defined set of validation rules on the data reported by their clients, and reject the reports, when these are not met.[14] Furthermore, in 2014, ESMA published "Questions and Answers on EMIR implementation" (EMIR Q&A), a regularly updated document, laying out guidance for interpreting the EMIR provisions. Finally, on 1 November 2017 the revised EMIR technical standards[15] entered into force, providing further clarity on reporting data on derivatives. A very important element of these reporting requirements was the obligation for the TRs to provide data to authorities in the format of standardized XML messages, compliant with the ISO20022 standard. This greatly facilitated the aggregation of TR information, fostered the development of dedicated IT infrastructures, and lowered the barriers of entry for analysts working on the EMIR data.

Another development contributing to improve the data quality was the EU-wide cooperation among authorities on the EMIR-related analysis, and corresponding knowledge sharing. An important initiative to note is the ESMA data quality framework, which enables the authorities to report to ESMA data quality or reporting issues. These issues are then prioritised and distributed by ESMA among the national supervisory agencies (NCAs – National Competent Authorities). Notwithstanding, the EU EMIR supervisory framework remains very complex, with 27 independent NCAs supervising their reporting entities, ESMA supervising the TRs, and over 100 authorities overall using the information provided under EMIR. The complexity of the framework leads to long feedback loops and limited capacity of the data quality assurance process.

[11] See Fache Rousová, et al. (2015)

[12] The need of providing clarifications have been addressed over time through regular updates of the EMIR Q&A on Implementation of the Regulation (EU) No 648/2012 on OTC derivatives, central counterparties and trade repositories (EMIR)

[13] See ESMA (2019) and ESRB (2022)

[14] To be compliant with the requirements of Article 19 of the Commission Delegated Regulation (EU) 150/2013, TRs should reject the reports which are not submitted in line with the reporting requirements specified in the *Validations* table.

[15] See https://www.esma.europa.eu/data-reporting/emir-reporting

Hence, the quality of information reported under EMIR is still far from satisfactory, and the respective analysis requires spending a considerable amount of time on data cleaning. Some data quality issues are easily identified by comparing aggregate statistics with other available benchmarks, e.g. information reported by BIS[16] under their semi-annual and triennial surveys, or by applying straightforward thresholds or plausibility check to the reported data. Other issues or anomalies are not easy to identify or interpret. While the quality issues are very heterogenous, the following main categories can be enumerated:[17]

1. The quality of the reporting is very uneven across the counterparties, i.e., some entities (including big players) report data of very low quality.
2. Reported information often does not match between the two counterparties reporting the same trade.
3. Many numerical values, including notional, contract value, and margins are highly implausible.
4. The direction of exposures is not reported correctly.
5. The quality of the collateral-related variables is low.
6. A large number of trades is not reported.[18]
7. Trades are not properly terminated by the reporting entities, and remain on the reports, even if they are not outstanding anymore.
8. Old outstanding legacy trades (reported before the introduction of TR validations, or before November 2017 update of technical standards) have inferior quality to the more recent ones.
9. Certain errors are introduced in the process of TRs' transformation of the received data.

2. ECB approach to cleaning and processing EMIR data

2.1. EMIR dedicated IT infrastructure

In 2017-2018 the ECB, jointly with the ESRB Secretariat, built a dedicated IT infrastructure to collect and process EMIR data. The EMIR IT system (Figure 1) collects the daily information from the ESMA TRACE platform,[19] and applies a series of transformations, including format conversions, enrichment with reference databases, data quality checks, and de-duplication of double-sided trades. The data is then made available to internal ECB and ESRB Secretariat users via the ECB analytical platform DISC.[20] Under the EMIR rules, the financial markets participants are obliged to report to TRs by T+1 the details of derivatives transactions traded on date T. The data collected by TRs, together with the additional reports compiled by them, are then made available to the authorities, either directly or via the ESMA TRACE portal, which acts as a central point of collection. Usually, the collection of data for the reference period T is completed by T+2 and data are processed by the IT system on T+3.

[16] Bank for International Settlements
[17] See also ESMA (2023)
[18] This phenomenon is difficult to measure in its entirety, but it can be corroborated by the fact that many trades are reported only by one counterparty, even if it is evident from the information reported that the other counterparty is also subject to mandatory reporting.
[19] The ESMA IT infrastructure through which EMIR data are collected by TRs and made available to authorities.
[20] The DISC Data Platform is a big-data platform (Hadoop) introduced in 2017 at the ECB, which provides a central, secure place for collecting, preparing, organising, storing and analysing data.

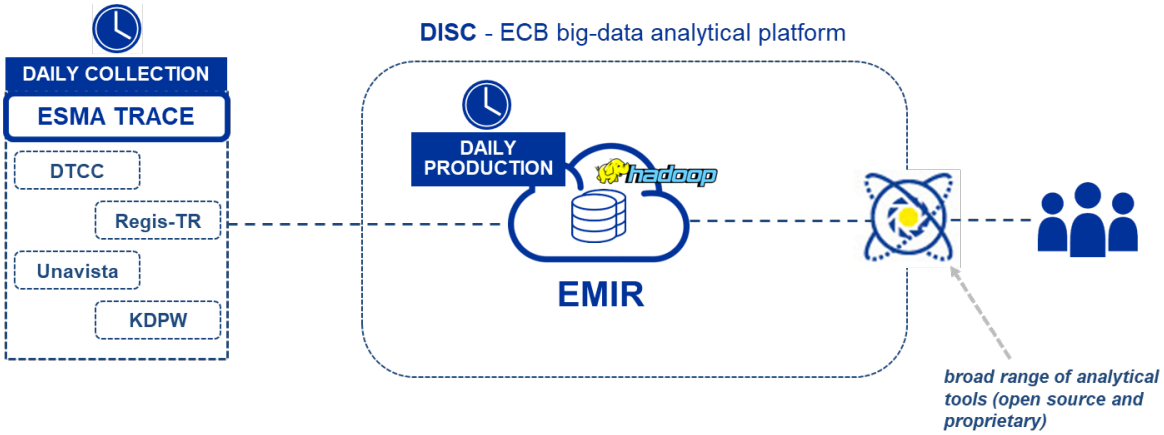

Figure 1: High-level diagram of the EMIR IT system

The EMIR dataset includes data on derivatives transactions reported by financial markets participants as well as reports compiled by the TRs, which are required to generate the trade state reports with transaction-level information on the stock of all outstanding contracts at a given date. TRs are also responsible for calculating aggregated positions by class of derivative and by reporting entity based on the details of the derivative contracts reported. In addition, TRs make available to authorities the relevant details of derivative reports rejected, and the reconciliation status of all derivatives reported for which the trade repository has carried out the reconciliation process.[21]

Reference databases available at the ECB or publicly available (e.g. GLEIF[22] or ISO10383[23]) are used to enrich the EMIR data with additional information on the basis of common identifiers for fields such as counterparties, underlying instruments, benchmark rates and foreign exchange rates. Furthermore, the tools provided by the DISC analytical platform allow the users to easily join the EMIR data with other granular information stored in DISC, for instance information on the underlying of the derivatives contracts identified through an ISIN, or the reference entity of a credit default swap identified with the Legal Entity Identifier.

The execution of data quality checks is an essential step of the data processing within the EMIR IT system. They are defined following the validation rules provided by ESMA,[24] and specify the checks that the TRs are expected to carry out when receiving the reports from the reporting entities. The validation process is performed at trade level and results in 139 quality flags added to the EMIR granular dataset providing users with a set of Boolean variables indicating, for instance, the compliance of the format of the LEIs identifying the counterparties involved in a trade or the completeness of the information in relation with the different action and observation types. The results of the quality flags are aggregated in a unique indicator to measure the data quality of the information reported for a trade. As additional data quality measures, the system calculates 20 data quality indicators computing statistics on the quality flags broken down by TRs, asset class, execution venue, observation type and reference period.

One of the distinguishing features of the EMIR reporting regime is the so-called "dual-sided reporting obligation"; that is, both counterparties of a trade are required to report all the details of that trade to a TR, and the ECB has access to the two reports if both counterparties are subject to mandatory reporting (e.g. when both counterparties are domiciled in the euro area). To facilitate the analysis by

[21] Trade repositories carry out a regular reconciliation process, identifying and comparing the two legs of the same trade.

[22] GLEIF – Global Legal Entity Identifier Foundation – publishes daily list of Legal Entity Identifiers and corresponding reference data, see: https://www.gleif.org/

[23] Market Identifier Codes, see: https://www.iso20022.org/market-identifier-codes

[24] See https://www.esma.europa.eu/sites/default/files/library/esma74-362-853_emir_validation_rules_for_revised_rts_its_0.xlsx

final users and to avoid the risk of double counting, an automated procedure for de-duplicating trades has been implemented within the EMIR IT system. Trades that can be paired, along with non-paired trades (i.e. the trade with only one side reported), are converted into a consistent format in order to build a new dataset in which every row has two separate sets of variables representing the information reported by the two counterparties of the contract. When the two sides of a trade can be found, the system calculates a matching score and determines the information that is likely to be the most reliable from the two sets of reported trade characteristics and stores it into a set of "best leg" variables.

2.2. Monitoring of collection, processing, and data quality

The timely provision of the EMIR data to users entails the monitoring of the collection of the EMIR data and of their processing. A set of tools has been developed to support the daily monitoring activities and to facilitate the detection of issues that can arise.

A report is produced daily in a fully automated manner with different sections that provide key pieces of information for monitoring the EMIR dataset. The completeness of the data is monitored checking that all reports expected by reference date, trade repository, and report type are successfully collected by the EMIR IT system. The report also includes detailed information for the reference dates processed on the outliers and on the transformation processes, such as the enrichment and the de-duplication.

The status of each reference date is presented to the users in a dashboard in a calendar format. It allows distinguishing among complete reference dates for which all data were received and processed successfully in the EMIR IT system and the ones that still have outstanding issues (red), e.g. the enrichment process failed. The source data of the calendar dashboard are available to users in a tabular format helping them to automate their analysis, e.g. programmatically excluding the incomplete dates and considering only the complete ones.

The analysis of data quality is not trivial for such complex dataset. In this respect, the ECB DG-S and ESRB Secretariat developed a framework, called Automated Data Quality (ADQ),[25] that allows to identify and classify the developments of numerical measures of granular and multi-dimensional datasets. The tool is based on binary trees that aim at identifying the relevant dimensions that are driving the changes observed in the data for a given measure. Several analyses can be run through the tool considering different quantitative measures (e.g. notional and contract value) and differentiating between time series or intraday (double sided) analysis. The kind of analysis is defined by the initial set of parameters that is taken as input by the ADQ process. The results of each analysis are summarized in a report but are also stored in a database.

2.3. DG-MF's EMIR cleaning procedures

DG-Macroprudential Policy & Financial Stability (DG-MF) staff has developed a cleaning procedure to further process the data and make them ready for analytical purposes. The procedure relies on several SQL queries and Python scripts that are executed in sequence. Each step creates an intermediate table, which feeds in the following step to create, at the end, one final cleaned table. The procedure cleans one date at a time and relies on the EMIR de-duplicated trade state table provided by DG-S EMIR team (as described in chapter 2.1). The scripts can be grouped in three categories: i) Basic Cleaning, ii) Segments, and iii) Margins. All scripts are connected in a modular structure, with specific segments dedicated to the following asset classes: interest rate derivatives (IRD), currency derivatives, equity derivatives and credit default swaps (CDS)[26]. The combination of all the modules yields a final table with de-duplicated EMIR data, enhanced with new fields created by the cleaning procedure.

[25] See also Agostoni, et al. (2023)

[26] The modules for the missing asset classes are currently under development.

2.3.1. Basic Cleaning

This script can be considered the core of the cleaning and is not asset-class specific. There are three queries, each building on the previous one, which create three intermediate tables. Each table enhance EMIR data with additional fields, tackling different data quality issues in the original reporting.

The first query leverages on the Classification of Financial Instrument code (CFI)[27] to clean the asset class and the contract type fields. When the CFI code is not available, the query looks at the reported asset class and contract type. It also creates a CFI consistency flag, that allows to identify those cases where the reported asset class and contract type are not consistent with the reported CFI code.

The second query tackles misreporting in the execution timestamp and fixes mis-reported maturities. It also defines several flags to identify potentially ambiguous trades, including whether the counterparty side (buyer or seller) is consistent between the two reports, whether a transaction is intragroup and whether it is centrally cleared. Then, it checks that the notional for each transaction is within the min-max range of its asset class. Such range is specified by the user.

The third query adds further quality flags, such as flags on maturity or on contract value.

2.3.2. Segments

The output table resulting from the basic cleaning queries is the input for the next steps in the procedure, where dedicated scripts and queries take care of cleaning specific asset classes and contract types, as follows:

- Interest rate derivatives: a function takes care of generating the different tenor curves and map them to EMIR data. The query then checks the consistency of the dates reported for an interest rate derivative. Finally, it creates a new field with a coarse taxonomy of product for the whole interest rate derivatives market.
- Currency derivatives: the main output of the script is producing a field with a coarse taxonomy of product for the currency forwards' segment. The focus is on the most traded products, grouping the others in macro categories to allow for meaningful statistics.
- Equity derivatives: a function maps different ISINs or proprietary codes, which refer to major indices, to the ISIN of the underlying itself. The scripts also classify options into European and American ones.
- Credit Default Swaps: the query further cleans the notional of CDS contracts (on top of the cleaning procedure in the Basic Cleaning) and creates ad-hoc flags to account for inconsistencies. In a second step, it distinguishes between single name (ISIN), index and other multi-name CDS (typically basket). In order to map the index and single name CDS, it merges with data from Datastream.

The segments are usually run one after the other, but they could also be run individually, according to the specific analytical need. The segment cleaning allows to better identify the type and characteristics of each derivative, much more in depth than how it is done with the basic cleaning. This becomes particularly relevant for certain types of derivatives, like interest rate swaps, that are largely used and have important financial stability implications (Section 3.2).

[27] ISO 10962, see: https://www.iso.org/obp/ui/#iso:std:iso:10962:ed-5:v1:en

2.3.3. Margins

The final step in the DG-MF cleaning procedure looks at margin data. The purpose of this module is to flag and identify potential inconsistent or misreported fields related to collateral on trade or portfolio level. The first step is identifying and reconciling the collateralisation level reported by the counterparties. There are four possible values the collateralisation field can take: i) fully collateralised; ii) partially collateralised; iii) one way collateralised; iv) uncollateralised. The code makes sure that the value reported in this fields is consistent with the margins that are reported[28] and with what both counterparties of the trade reported[29]. Then, the code defines new identifiers for the collateral portfolio codes, so that all portfolios have one, even those where such code is not reported. The new identifiers are built in a way so that each portfolio is uniquely identified from the perspective of the reporting counterparty.

2.3.4. Module for collateralised portfolios

In addition to the cleaning procedure described above, two additional modules create separate initial and variation margins tables, using as input the final cleaned table generated by the cleaning procedure described in the previous paragraphs. These modules are an outcome of joint DG-MF and DG-Market Infrastructure & Payments (DG-MIP) work on initial margin calculations conducted in 2020 and 2021, with focus on analysis of the March 2020 market turmoil. By producing portfolio level data (rather than transaction-by-transaction data) and not keeping the de-duplicated structure of the data, the module follows a different approach from the rest of the MF EMIR cleaning infrastructure. This approach helps overcome the inconsistencies that arise if two counterparties report their common portfolios differently.

3. Data application for financial stability analyses

Together with other database available at the ECB, EMIR tables, cleaned with the procedure described in the previous section, have been used as inputs to financial stability assessments on an ad-hoc basis or for monitoring purposes. The range of applications is very broad and covers multiple asset classes and sectors. For example, it might be related to a specific entity (e.g. when the analysis is limited to the outstanding derivatives positions of a single institution), sector (e.g. when during the Covid-19 period some sectors have been affected by liquidity strains with implications for margin calls), or historical episodes (e.g. like the recent geopolitical events). However, these analyses present some caveats which limit the reading of the results. EMIR reporting suffers from some limitations, for example related to the limited geographic coverage – which does not include reporting by non-euro area subsidiaries of euro area reporting counterparties, constraints to consolidation, etc) which complicate analysis. Two examples follow in the section.

3.1. Commodity derivatives trading in the wake of the Russian invasion of Ukraine

Energy-related commodity prices and volatility started rising in mid-2021 and reached unprecedented heights in March and during the summer of 2022 following the Russian invasion of

[28] For example, if the reporting counterparty indicates that the contract is partially collateralised then variation margin (VM) posted should be populated, initial margin (IM) posted should be null or 0 and VM and IM received should be null, 0 or populated.

[29] For example, if one entity reported the trade as fully collateralised, the other counterparty cannot report it as uncollateralised.

Ukraine. The extreme price movements that occurred at that time highlight the importance of energy derivatives markets for hedging risks in the energy sector, as well as some of the pressures that can arise in these markets.

In a Special Feature published in the November 2022 ECB Financial Stability Review (FSR)[30] , authors provide an overview of the European energy derivatives market, with a focus on natural gas and power. It analyses the impact of extreme energy prices on the structure of energy markets, the liquidity stress faced by entities with the largest exposures to market risk, and the risks that their vulnerabilities may pose to their counterparties in derivatives and credit markets.

To run the analysis authors largely relied on derivatives data reported under EMIR, and on some generic fields as well as the specific fields referring to the asset class *Commodity.* For example, among the generic fields some of them are pivotal: LEIs of the counterparties, the buyer/seller side of each trade, the notional, the clearing and intragroup flags, the execution and the maturity dates, and the currency. These fields usually are the one resulting from the application of the EMIR code repository described in Section 2.3. To identify the country and the sector of the counterparties the LEIs have been matched with the database produced along the procedure described in Lenoci and Letizia (2021), while euro area banks have been consolidated according to the group structure reported by the Single Supervisory Mechanism (SSM). The most important fields referring to *Commodity* derivatives and used to identify contracts having as underlying natural gas are the commodity base, the commodity underlying and the ISIN of the underlying product (which is also matched with commercial data providers to confirm the reporting).

3.2. Interest rate swaps trading in the context of rising rates

Interest rate derivatives are a key instrument for risk management. Within this asset class, interest rate swaps (IRS) and forward rate agreements (FRAs) represent the most traded contract types covering 72% of the overall derivatives market. Trading of euro-denominated IRS and FRAs has risen sharply since 2021, reflecting the critical role of derivatives in managing interest rate risk due to the shift in monetary policy expectations. Banks trade a large share of interest rate derivatives, as these contracts are an essential tool to hedge, speculate and manage risks.

Due to the prominent role played by IRS and to its large use by the financial sector, ECB staff worked extensively on the development of a code to clean the fields reported under EMIR which are specific for these derivatives contracts, as described in Section 2.3.2For example, for each of the two reporting legs, the effective date of the trade, the execution timestamp, and the settlement date - when available - have been used to identify the starting date of the swap contract. The identification of the starting date of the swap contract, together with the contractual maturity of the swap have played a pivotal role to match banks' maturity profile of balance sheet items (e.g. loans and securities) with the maturity of derivatives contracts. Along the same line, the reported floating rates, the product classification code, and the fixation period have been used to identify the reference rate of the floating leg. Moreover, these variables allow the identification of contracts falling within the perimeter of swaps having as underlying the Euribor rate.

Financial stability analyses using EMIR data on interest rate derivatives revealed details about positioning of market participants which are consistent with their risk management needs. [31] Three groups of euro area banks participate in the swap market. Banks which benefit from higher interest rates tend to take on offsetting positions which reduce the volatility of their profits and capital. The same holds for banks benefitting from lower interest rates. Insurance companies and pension funds are the main providers of interest rate hedges, positioning themselves to receive fixed-rate payments for long-term maturities, in line with their aggregate negative duration gap. At the centre of the market, large euro area banks play a market-making role, facilitating transactions and accounting for a large share of all trades, but not taking substantial directional positions. The large

[30] See Furtuna, et al. (2022)

[31] See Dries, et al. (2022)

scale of interest rate derivative trading exposes participants to potentially sizeable margin calls: for example, a parallel shift of 100 basis points in the yield curve would lead to a wealth transfer (equivalent to a margin payment) of around € 90 billion in aggregate terms as of June 2022. .

4. Conclusions

As illustrated in this paper, the EMIR data is a valuable source of information on the derivative contracts executed in the European Union, with a great potential for financial stability analysis by the EU authorities. Nevertheless, the complexity and granularity of the framework still present considerable challenges for the reporting agents and the supervisors. As a result, data quality is far from satisfactory and a significant effort in preparing and cleaning the data is required to draw accurate conclusions from the information received.

In this context, the crucial component is having structured and regular processes with regards to the data, supported by adequate IT infrastructure and tools, as well as appropriate competences and expertise within the institution. Such processes involve in-depth monitoring of data collection and processing, analysis of data quality, and replicable data cleaning procedures. Such approach leads to synergies at the level of the institution and beyond, reducing the effort required to start working with the data.

The data could also be further improved by promoting the use of standards in reporting, issuing more definitive guidelines and reducing room for interpretation. On 20 May 2019 the EU Commission implemented a major amendment of the EMIR Regulation, known as "EMIR Refit",[32] aiming at enhancing and streamlining the reporting of derivatives in the EU. This was followed by the publication of updated EMIR Technical Standards, reporting instructions, and detailed reporting guidelines, applicable as of 29 April 2024.[33]

The new framework follows the globally agreed guidelines on Unique Transactions Identifier (UTI), Unique Product Identifier (UPI) and other Critical Data Elements (CDE).[34] Furthermore, the new legislation mandates a common reporting format, in line with ISO20022 standard, for the information reported to trade repositories. While the scope of the changes may lead to some temporary disruptions after the go-live, it will translate to better quality and reliability of the data in the medium term.

Finally, the cooperation between different authorities is another way to realize synergies in the use of this information, e.g. for macroprudential analysis. While such cooperation is already on-going on some international fora, it is often hampered by legal obstacles to sharing information, and related uncertainty. This applies both to sharing within EU jurisdiction, as well as with other non-EU authorities. Improving the legal framework for sharing granular information on derivatives could lead to streamlining the global cooperation in this respect.

In conclusion, EMIR data constitutes an essential source of information on derivative markets for EU authorities. Given its volume and deficiencies of the data stemming from inaccurate reporting, efficient use of this data requires development of robust processes around data. The ECB is continuously extending and refining its tools to further benefit from this important resource.

---

[32] See Regulation (EU) 2019/834 of the European Parliament and of the Council of 20 May 2019 amending Regulation (EU) No 648/2012 as regards the clearing obligation, the suspension of the clearing obligation, the reporting requirements, the risk-mitigation techniques for OTC derivative contracts not cleared by a central counterparty, the registration and supervision of trade repositories and the requirements for trade repositories

[33] See:
- https://www.esma.europa.eu/data-reporting/emir-reporting
- https://www.esma.europa.eu/press-news/esma-news/esma-publishes-guidelines-and-technical-documentation-reporting-under-emir

[34] See: https://www.leiroc.org/international_bodies.htm

References:

G. Agostoni, L. de Charsonville, M. D'Errico, C. Leonte, G. Skrzypczynski (2023), "Anomaly intersection: disentangling data quality and financial (stability) developments in a scalable way", https://www.bis.org/ifc/publ/ifcb59_32_rh.pdf

T. Carraro, L. Fache Rousová, O. Furtuna, M. Ghio, K. Kallage, C. O'Donnell, F. Vacirca, S. Zema (2021), "Lessons learned from initial margin calls during the March 2020 market turmoil", https://www.ecb.europa.eu/pub/financial-stability/fsr/html/ecb.fsr202111~8b0aebc817.en.html#toc40

ESMA (2023), "2022 Report on Quality and Use of Transaction Data", https://www.esma.europa.eu/sites/default/files/2023-04/ESMA74-427-719_2022_Report_on_Quality_and_Use_of_Transaction_Data.pdf

L. Fache Rousová, M. Osiewicz, G. Skrzypczynski (2015), "The challenges of standardization and aggregation of EMIR data in Europe: Six trade repositories and 28 countries", https://www.bis.org/ifc/events/ifc_isi_2015/089_rousova_osiewicz_skrzypczynski_paper.pdf

FSB (2022), "OTC Derivatives Market Reforms – Implementation progress in 2022", https://www.fsb.org/wp-content/uploads/P071122.pdf

Dries J., Klaus B., Lenoci F., and Pancaro C. (2022) "Interest rate risk exposures and hedging of euro area banks' banking books" ECB Financial Stability Review, May 2022

Grassi G., Kockerols T., Lenoci F., and Pancaro C. (2022) "Euro area interest rate swaps market and risk-sharing across sectors" ECB Financial Stability Review, November 2022

O. Furtuna, A. Grassi, A. Ianiro, K. Kallage, R. Koci, F. Lenoci, A. Sowiński, F. Vacirca (2022), "Financial stability risks from energy derivatives markets", https://www.ecb.europa.eu/pub/financial-stability/fsr/html/ecb.fsr202211~6383d08c21.en.html#toc38

F. Lenoci, E. Letizia (2021), "Classifying Counterparty Sector in EMIR Data", https://link.springer.com/chapter/10.1007/978-3-030-66891-4_6

ESMA (2019), "Final Report - Peer review into supervisory actions aiming at enhancing the quality of data reported under EMIR", https://www.esma.europa.eu/sites/default/files/library/esma42-111-4895_emir_data_quality_peer_review.pdf

ESRB (2022), "ESRB's view regarding data quality issues and risks for financial stability", https://www.esrb.europa.eu/pub/pdf/other/esrb.letter220713_on_data_quality_issues~18eccb6993.en.pdf