BANCA D'ITALIA
EUROSISTEMA

EMIR data analytics for research, financial stability and supervision

Seminari e convegni
Workshops and Conferences

number 27

July 2025

BANCA D'ITALIA
EUROSISTEMA

EMIR data analytics for research, financial stability and supervision

Seminari e convegni
Workshops and Conferences

Il testo è disponibile nel sito internet: www.bancaditalia.it

ISSN 2281-4337 (print)
ISSN 2281-4345 (online)

Grafica a cura della Divisione Editoria e stampa della Banca d'Italia

This is a collection of papers that were presented at the Banca d'Italia workshop on *"EMIR data analytics for research, financial stability and supervision"*, held in Rome on 6 and 7 June 2024 and organized in collaboration with the European Systemic Risk Board. The workshop was organized by Michele Leonardo Bianchi, Francesco Franceschi, Dario Ruzzi, Marco Romani and Roberto Stok[1] (Banca d'Italia) and Marco D'Errico (European Systemic Risk Board Secretariat). The papers included in this collection have been edited by Michele Leonardo Bianchi, Francesco Franceschi, Dario Ruzzi, and Gwyneth Schaefer (Banca d'Italia), with the consent of the papers' authors.

1 European Systemic Risk Board Secretariat until June 2024.

Contents

Introduction **4**
by M.L.Bianchi, F.Franceschi, D.Ruzzi, and G. Schaefer (Banca d'Italia)

EMIR data quality – current state and way forward **7**
by G.Agostoni, N.Calabrese (ECB), M.D'Errico (ESRB), L.Henkel, and G.Skrzypczynski (ECB)

Assessing initial and variation margins reported under EMIR **18**
by A.Pioli and F.Vacirca (ECB)

How fair is the value of contract reported under EMIR? An analysis on interest rate derivatives **27**
by M.L.Bianchi and D.Ruzzi (Banca d'Italia)

EMIR article 9 data analytics software **40**
by G.Grigorescu (National Bank of Romania)

Implementation of EMIR Refit reporting at Banque de France: technical challenges and insurance supervision use case **48**
by F.Pégory and M.D'Ambrosio (ACPR)

Mining equity implied volatilities from EMIR **56**
by M.L.Bianchi (Banca d'Italia) and D.Miceli (ECB)

A holistic approach to EMIR data by the Central Bank of Ireland: Implementing the EMIR Roadmap **72**
by G.Beattie, A.Karapappas, S.Fragkoub, G.Insalaco, E.Flood, J.Harte, and D.Quinn (Central Bank of Ireland)

Leverage, liquidity and concentration: Vulnerabilities of GBP liability-driven investment (LDI) funds through the prism of regulatory data **86**
by A.Bouveret and H.Leclerc (ESMA)

New NBFI monitoring tool **103**
by A.Ianiro, L.Rousová (ECB), and D.Ruzzi (Banca d'Italia)

The effects of obligatory central clearing for Dutch pension funds **113**
by M.Driesse and E.van der Woerd (DNB)

Derivatives of Belgian banks and insurance companies: Three case studies into exposures, rising interest rates, and the energy crisis **122**
by S.B.Hadj, R.Faber, P.Rovira Kaltwasser, and P.Van Roy (NBB)

Luxembourg derivatives market 2023 – CSSF & CAA joint report on market structure, market trends and data quality **135**
by A.Gentilini (CSSF)

Introduction

Michele Leonardo Bianchi,[a,1] Francesco Franceschi,[a] Dario Ruzzi,[a] and Gwyneth Schaefer,[a]

[a]*Financial Stability Directorate, Banca d'Italia, Roma, Italy*

This volume is a collection of some of the papers presented at the workshop on *"EMIR data analytics for research, financial stability and supervision"* held at the Banca d'Italia in June 2024 and organized in collaboration with the European Systemic Risk Board (ESRB).[2]

In 2014, the European Market Infrastructure Regulation (EMIR) introduced a requirement for entities resident in the European Union (EU) to report, on a daily basis, detailed data on derivative transactions. Reports include information on the trade counterparties and on a large set of contract characteristics, such as notional amount, underlying asset and maturity date; they are submitted to trade repositories (TRs) and subsequently shared with competent EU authorities based on their mandates. The resulting database is commonly referred to as *EMIR data*.

Banca d'Italia and ESRB brought together data scientists, researchers, supervisors and policymakers sharing a common goal: to deepen the understanding and enhance the capabilities of EMIR data. Over one hundred registered participants, both in-person and remote, from approximately 40 institutions, attended the workshop. Their presence is testament to the importance of these topics and the commitment to them in day-to-day activities.

The aim of the workshop was to foster the sharing of information and lessons learned from the use of this complex dataset and to promote the dissemination of knowledge. Knowledge of EMIR data has a wide range of meanings: from understanding potential legal issues, such as data confidentiality, to completing highly technical onboarding forms for trade repository (TR) data retrieval, identifying possible data gaps, pricing complex derivative contracts, and analyzing margins, including collateral portfolio-trade relationships.

To fully understand EMIR data, one needs strong knowledge of derivative markets, advanced programming skills and, last but not least, the patience to read,

[1] The authors thank all speakers and participants, for contributing to the outstanding quality of the workshop. The views expressed are those of the authors and do not necessarily reflect those of Banca d'Italia.

[2] See https://www.bancaditalia.it/media/notizia/workshop-emir-data-analytics-for-research-financial-stability-and-supervision/.

understand, and interpret the regulations. In addition, there are data quality issues, including data gaps, related to the data collection and processing phases. Dealing with EMIR data is not a simple task. Additionally, the journey of reports from the reporting entity to the analyst (researcher or data scientist) is far from linear and asking the reporting counterparties for an explanation of a single submission can be challenging. Although much has been achieved with respect to the past, all the entities involved in the process (e.g., reporting counterparties, trade repositories, and competent authorities) still could and should do more to enhance the usability of this complex dataset.

The workshop was divided into three sessions:[3]

- *Advanced analytics.* In this session, the importance of data quality was explored from various perspectives. Ten years after the introduction of EMIR reporting, data quality is still an issue and much work is needed to go from raw data to a dataset ready for analysis and research. The speakers discussed these issues, both from a broad perspective and by focusing on specific products (futures and interest rate swaps) and data fields (margins and contract values).

- *Research and policy.* In this session, EMIR data met academic research. Among the presentations were: (1) a novel methodological framework that offers some advancement in measuring synthetic leverage; and (2) a framework to assess systemic risk related to the inadequate liquidity preparedness of non-bank financial institutions (NBFIs) to margin calls under adverse scenarios.

- *Risk monitoring and supervision.* This session focused on frameworks and tools for risk monitoring and supervision. Some presentations examined the monitoring of derivatives markets in specific countries, while others explored analyses of particular non-bank entities, such as investment funds and pension funds; still others addressed specific risks like interest rate risk.

As emphasised by several speakers during the workshop, the EMIR data quality process is challenging, mainly because the number of fields to check is large. Much has been done to reconcile EMIR with alternative sources (e.g. FINREP or market data). However, as long as we are unable to determine the fair value of the contract, the assessment of the risks associated with the derivative positions cannot be accurate. Contract value determination is certainly possible, at least for simple products (like interest rate swaps and forwards). At the Banca d'Italia we started to fully revaluate EMIR positions with two aims: (1) to conduct advanced data quality checks; and (2) to perform sensitivity analyses. In our view, the revaluation will be extremely useful for extracting risk indicators from this huge dataset.

EMIR data is a unique and rich database containing valuable information; being confidential, however, is accessible only to competent authorities. In order to push forward the research agenda based on EMIR data, collaboration with academic

[3] As more contributions than expected have been received, a Banca d'Italia colleague suggested to add a poster session to allow all submitters of a contribution to present their work. The posters allowed us to "visualize" five very interesting works on EMIR and discuss them.

researchers in quantitative finance, network analysis, and financial econometrics is very important.

In conclusion, a final technical comment. More granular and more timely data are being collected, and not only EMIR data but also other complex and granular datasets, such as security financing transaction (SFTR) data. Research or monitoring tools that combine these datasets, along with potentially others like security holding statistics (SHS), could significantly improve our understanding of the derivatives market and of the assets underlying these derivatives, as well as the systemic risk implications, and help to the development of macroprudential policies.

EMIR data quality – current state and way forward

Gemma Agostoni,[a,1] Nicola Calabrese,[a] Marco D'Errico,[b] Lukas Henkel,[a] and Grzegorz Skrzypczynski[a]

[a] *European Central Bank, Directorate General* Statistics
[b] *European Systemic Risk Board Secretariat*

Abstract. The European Market Infrastructure Regulation (EMIR), one of the European Union (EU) responses to the global financial crisis, introduced a range of measures aimed at achieving better functioning and transparency of the derivatives market. These measures include the obligation to report transaction-level derivatives data to trade repositories (TRs), which are then shared with over one hundred authorities in the EU.

Since 2014, these data have been intensively used by the European Central Bank (ECB) and European Systemic Risk Board (ESRB) Secretariat for a variety of tasks, including financial stability monitoring, central counterparties (CCP) oversight, and micro-prudential supervision. In 2024, a major overhaul of the reporting requirements, following the so-called EMIR Refit amendment, went live. This broadened the set of data available to regulators, also introducing clarifications of concepts and a more comprehensive data model to report information on derivative contracts. At the same time, the new reporting framework presents new challenges to TRs, reporting agents, and final users of the data, which will take time to resolve.

This short paper describes the EMIR dataset and how it is used by the ECB and ESRB Secretariat. It highlights the main data quality challenges encountered by data users and provides initial insights into the quality of information reported under EMIR Refit. Finally, it discusses several recommendations that could enhance data quality, some of which could be implemented with relatively limited effort from reporting entities, TRs, and regulators.

[1] The authors are grateful to Sébastien Pérez-Duarte and Javier Huerga Aramburu for their helpful suggestions. This paper should not be reported as representing the views of the European Central Bank (ECB) nor of the European Systemic Risk Board (ESRB) or its member institutions. The views expressed are those of the authors and do not necessarily reflect those of the ECB nor the official stance of the ESRB or its member institutions.

Key words: EMIR, Refit, derivatives, data quality, granular data.

1 Introduction

In 2009, as a response to the global financial crisis, G20 leaders agreed to introduce changes to increase the efficiency and transparency of markets, including those for derivatives. The European Union (EU) legislation aimed at making derivatives markets safer was called the European Market Infrastructure Regulation (EMIR), and introduced a range of measures, including the mandatory clearing of standardised derivatives, risk mitigation techniques for non-cleared trades and requirements to monitor and mitigate the operational risks associated with derivatives. Another significant pillar of EMIR, introduced in February 2014, was the obligation for counterparties to report detailed and daily information regarding derivative transactions. This information is reported to trade repositories (TRs) and is then shared with over one hundred authorities in the EU, depending on their mandate. A significant subset of this information is shared daily with the ECB and the European Systemic Risk Board (ESRB) Secretariat.[2]

To enable efficient data collection and processing, the ECB, in collaboration with the ESRB Secretariat, built the EMIR IT system which is a dedicated infrastructure specifically designed to handle this extensive and complex dataset. The system has been integrated into the ECB's big-data analytical platform and allows analysts to use a range of tools to retrieve insights from EMIR data in an efficient manner. Since its inception in 2017, the EMIR IT service has been instrumental in managing this immense volume of data, processing over 80 million observations daily. The IT system supports more than one hundred individual users at both the ECB and the ESRB Secretariat. These users leverage the processed data for a wide range of ECB tasks, including financial stability monitoring, central counterparties (CCP) oversight, and micro-prudential supervision.[3]

Over time, the ECB and the ESRB Secretariat have developed a comprehensive suite of daily processes and tools to effectively monitor the dataset generated by the EMIR IT system. With a dataset of such size and complexity, it is not feasible to employ traditional manual or semi-manual techniques for monitoring and data quality assurance.[4] A small, dedicated team, with the support of highly automated tools, carries out a daily process to ensure that the data available to the analysts are as complete and accurate as possible.

2 EMIR data quality

From the beginning of the data collection, EMIR data faced a large number of data quality issues, making it difficult for analysts to use this information efficiently for the exercise of their tasks. Significant effort is required to understand, clean, and

[2] The ECB has access to the contracts linked to the euro area, the ESRB Secretariat can access the full EU dataset. See Commission Delegated Regulation (EU) 2022/1856.

[3] See Agostoni et al. [2024b].

[4] See Agostoni et al. [2023].

aggregate this information. The complex supervisory framework[5] leads to frequent bottlenecks and long feedback loops in the data quality assurance process.

While the range of possible data quality problems is very broad, the main challenges in using the data can be classified into the following categories:[6]

- implausible numerical values;
- inconsistent and inaccurate reporting of margins;
- misreporting the direction of exposure;
- errors in trade lifecycle reporting;
- issues stemming from delegation of reporting;
- lack of reconciliation between counterparties;
- misreporting of timestamps;
- IT bugs in TR systems;
- underreporting.[7]

At the same time, it is important to note the variety of measures implemented by ESMA and other EU authorities to improve the quality since the beginning of reporting. These measures are illustrated in Figure 1 and include, among others:

- validation and reconciliation processes executed by TRs;
- development of the ESMA data quality framework;
- provision of guidance in the form of guidelines and Q&As;
- revision of legal texts to provide clarity and extend the data model (in 2017 and 2024).

The most notable revision is the so-called EMIR Refit amendment, which resulted in a major overhaul of the reporting framework in April 2024.[8] An important development that had a negative impact on the quality of data was the UK leaving the EU; this limited access to information reported by UK entities on EU-UK trades, which had been used to compare the information submitted by the two reporting agents and allowed users to spot discrepancies in reporting. Given the importance of UK entities, including some of the major central counterparties

[5] Trade repositories collect and process the data, over one hundred authorities access the data, including 28 relevant supervisory agencies, i.e., 27 national competent authorities (NCAs) and European Securities and Markets Agency (ESMA).

[6] See also Agostoni et al. [2024a] for more discussion on the taxonomy of data quality issues.

[7] The EMIR reporting obligation is double-sided, meaning that both counterparties to the trade have to report the details of the trade, as long as they are legal entities from the EU. The information reported under EMIR allows the regulators to identify entities that are obliged to report but have not done so. See also Perez-Duarte and Skrzypczynski [2019] and ESMA [2024].

[8] See chapter 3 for details.

(CCPs), this had been an important validation technique applied by users to clean the data.

Importantly, the dialogue and sharing of experiences between various EU authorities have increased over the years, facilitating work on this dataset and allowing regulators to draw important insights, relevant for their policy work. Nevertheless, the quality of EMIR data remains one of the biggest obstacles in using this information efficiently and at scale.

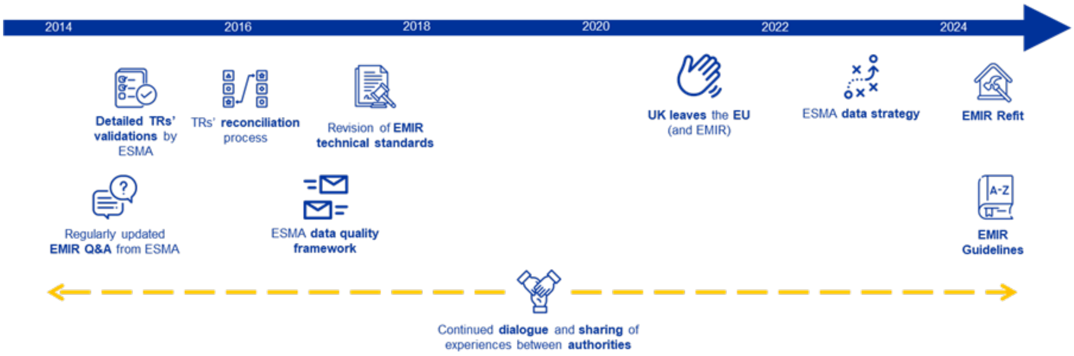


Figure 1 – The evolution of data quality measures in the EU. Source: ESMA [2022b], ESMA [2022a], ESMA [2024], EMIR Level 1 and Level 2 legislation.

3 EMIR Refit - first insights

In 2015, the European Commission initiated a comprehensive review of the EU regulations with the aim of making the EU laws simpler, more targeted, and easier to comply with.[9] One of the regulations that was thoroughly reviewed was EMIR. This led to the publication of the EMIR Refit amendment on 20 May 2019, which introduced significant modifications aimed at streamlining and enhancing the efficiency of the EMIR framework. Following the EMIR Refit, a process was undertaken to revamp the related Level 2 legislation concerning EMIR reporting rules. This was aligned with global efforts towards the harmonisation of OTC derivatives reporting, coordinated by international regulatory bodies, including the Committee on Payments and Market Infrastructures (CPMI), the International Organization of Securities Commissions (IOSCO), the Financial Stability Board (FSB), the Regulatory Oversight Committee (ROC).[10]

EMIR Refit introduces major changes to the reporting framework, including:

- increase of the number of reporting fields (from 129 to 203);
- change of the schema of reports submitted to TRs (from proprietary formats to ISO20022 standard);

[9] See https://commission.europa.eu/law/law-making-process/evaluating-and-improving-existing-laws/Refit-making-eu-law-simpler-less-costly-and-future-proof_en.

[10] For more information on global standardisation initiatives for derivatives see: https://www.leiroc.org/international_bodies.htm.

- Additional data quality measures (TR validations and reconciliation defined in Level 2 legislation, TR feedback to reporters);
- comprehensive model to report the lifecycle of trades, including linking identifiers;
- alignment with global standards on UPI, UTI, and CDE.[11]

These measures are expected to be an important step towards better data quality and increased use of this information among authorities. From our initial investigation, it appears that the change has proven to be challenging for reporting agents, TRs, and some regulators, and considerable effort will be needed to reach and subsequently surpass the quality of information reported under the old framework. Below, we have listed some initial insights into data reported under EMIR Refit.

3.1 Delivery of reports by TRs

In the initial days following the go-live of EMIR Refit on 29 April 2024, TRs encountered several issues that hindered the smooth delivery of reports. While some TRs were able to address and resolve these issues swiftly, others continued to face challenges, some of which are still ongoing. These challenges included, on one hand, missing or schema-incompliant reports, as well as late delivery. On the other hand, the reports conveyed inaccurate information, ranging from missing or duplicated transactions to data that was inaccurately or untimely updated. While most critical issues are believed to be resolved, the re-generation of missing or incorrect reports from the period April to September 2024 was not yet concluded as of October 2024.

Another problem was the differences in methods applied by the TRs to compile information transmitted to authorities. This is particularly important in the cases of reports derived by the TRs, like the so-called "state" reports.[12] In particular, we have observed varying approaches to:

- the construction of the margin state reports, and
- the adaptation of information reported before Refit to the new report structure.

These examples suggest that more detailed guidance would be beneficial for the TRs on how to unambiguously prepare reports for the authorities.[13]

[11] Unique product identifier (UPI), unique transaction identifier (UTI), and critical data elements (CDE).

[12] EMIR Refit foresees two types of state reports: trade state reports and margin state reports. Unlike "activity" reports, they present the current status of all outstanding derivatives at the end of the reference date of the report and are compiled on a daily basis by TRs from information received from reporting agents.

[13] See Section 4 for further discussion.

3.2 Number of observations and reporting population

The number of observations available in the ECB reports fluctuated significantly in the first few months of EMIR Refit, mainly due to fixes applied by the TRs to the reports shared with authorities. In September, the volume of the reports started to stabilise. A comparison with pre-Refit data reveals that the number of reported transactions has not significantly changed. As can be seen in Figure 2, the number of outstanding trades in the trade state reports increased slightly, which may simply reflect the developments in the euro area derivatives markets. Based on the evolution of the number of trades, it seems unlikely that some relevant entities stopped reporting after the EMIR Refit go-live. The reporting population has not changed significantly, with the vast majority of entities reporting before Refit continuing to report after the Refit go-live.

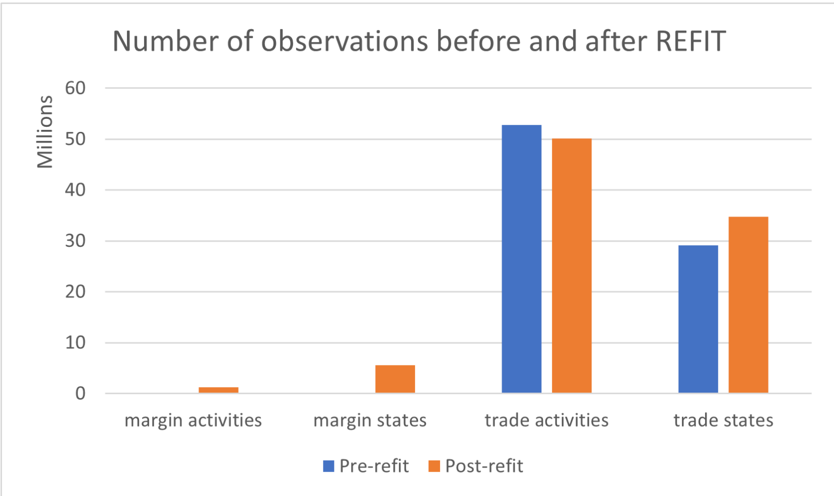


Figure 2 – Number of observations in ECB EMIR reports before and after Refit. Average number of trades in daily reports in the weeks of 15-19 April 2024 (before Refit) and 7-11 October 2024 (after Refit), own calculations. Source: EMIR data.

3.3 Implausible numerical values

Implausible numerical values in EMIR have always been rather common, and this has not changed with EMIR Refit. The aggregation of the notional amounts or contract values usually requires a considerable amount of time to identify and clean values that do not reflect the economic reality of the underlying contracts.

One notable pattern in EMIR Refit data is the presence of implausible notional values that amount to 9,999,999,999,999,999,999,999,999. This specific value is allowed by ESMA guidelines serving as a placeholder when the actual value is not yet available. In practice, entities often use any number of 9s, either before or after the decimal point, to indicate a pending value. This practice complicates the identification of these provisional values, making it challenging to distinguish between actual implausible values and legitimate placeholders.

Figure 3 presents the evolution of extreme notional amounts (after removing the placeholder values). To mark a value as implausible, a very conservative threshold of EUR 1 trillion is applied. The chart indicates a significant increase in the number of such implausible values following the go-live of the EMIR Refit. Around the middle of September, however, the number of implausible values drops, settling somewhat below the pre-Refit level. This is certainly a positive development, reflecting the TRs' and data reporters' efforts to improve the quality of reporting.

The implausible values of such magnitude can be very easily identified. The task becomes much more complex when values are on the verge of plausibility, making it very difficult to determine whether the value represents an actual characteristic of a contract or a simple mistake. This may materially impede the analysis of the data and is often difficult to tackle using automated approaches.[14]

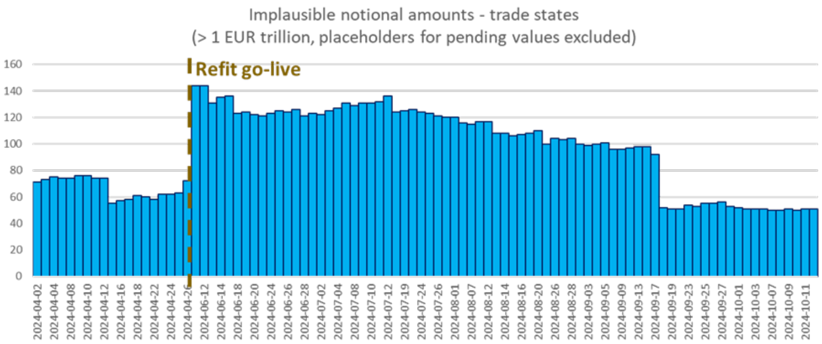


Figure 3 – Number of trades with implausible notional amount in the ECB trade states report. Source: EMIR data, ECB own calculations.

3.4 Reporting of the direction of exposure

One of the essential features of EMIR is the reporting of the daily valuation, allowing regulators to monitor the flow of risk in the economy. The reporting obligation under EMIR is double-sided, which means that a significant number of trades will appear twice in the dataset. For a given trade, the sign of the contract value reported by each counterparty should be opposite; if one party reports a long position, the other must report a short position. However, this fundamental principle has often been misreported, leading to significant difficulties in determining the direction of the exposure and impairing aggregation.

Several factors may contribute to these misreporting issues. One common reason is that a counterparty might receive the valuation from the other counterparty, such as a CCP, and fail to reverse the sign when submitting their report. Another frequent issue arises when the reporting obligation is delegated to one of the counterparties. In such cases, the delegated party might submit its own contract value for both reports, with the same sign, reflecting its own direction of exposure.

Yet another issue may be caused when completely different valuations are reported by both entities. While the valuation of the contract is supposed to be reconciled between the two counterparties, it is still often the case that the values differ significantly. One of the factors exacerbating this issue may be the new

[14] See also Agostoni et al. [2023].

EMIR guideline, mandating that the reporting agents report the contract value of the so called STM (settled-to-market) contracts[15] as a change from the previous day, instead of as a stock. As we can observe in the data, two counterparties to the same trade sometimes follow different approaches to valuation (stock vs. flow) making the contract value impossible to reconcile.

Figure 4 presents the share of outstanding trades affected by the above-mentioned issues. Since the EMIR Refit go-live, there has been a noticeable increase in the number of inconsistent valuation directions. Until September 2024, this could be partially explained by an IT issue at one of the trade repositories, which led to contract value data not being properly updated. Nevertheless, even in October data, less than two thirds of double-sided trades exhibited consistent valuations, which is noticeably below the pre-Refit levels.

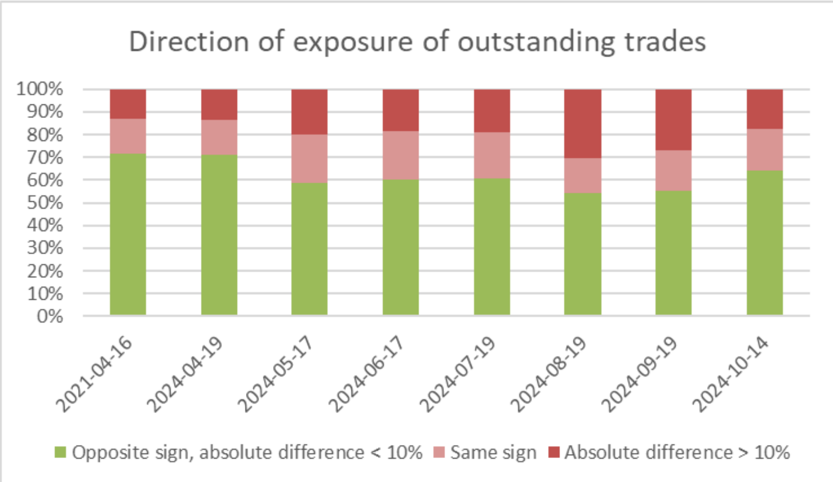


Figure 4 – Alignment of the direction of exposure between the counterparties to the trade. Source: EMIR data, ECB own calculations.

4 EMIR data quality – way forward

While significant steps have been made in the implementation of the EMIR Refit and the overall regulatory framework for OTC derivatives, persistent and pervasive data quality issues continue to pose a substantial obstacle to the efficient use of granular data on derivatives. These data quality challenges have far-reaching implications.

Firstly, low data quality increases opacity for policymakers and market participants. Inaccurate, incomplete, or inconsistent data obscures the true state of the derivatives market, making it difficult to gain a clear understanding of market dynamics. This opacity undermines the primary objective of EMIR, which is to enhance transparency and oversight in the derivatives market. Secondly, poor data quality hampers the scalability of monitoring frameworks for policymakers. When

15 The settled-to-market contracts reflect an accounting practice where the variation margin of a derivative is recorded as a settlement payment rather than as a transfer of collateral. See also: https://www.bis.org/press/p180920a.htm.

data quality is compromised, these frameworks struggle to scale up and adapt to the evolving complexities of the market. Finally, poor data quality increases uncertainty in analytical outcomes, making it difficult to distinguish whether an observed development reflects genuine market dynamics or is a result of inaccurate reporting.

The problem of data quality is a multifaceted one and may have different root causes. In this environment, it is crucial to focus on the elements that promise the greatest improvement with limited effort required from reporting agents, TRs, and authorities. Below, we propose a few recommendations that, in our view, could significantly increase the quality of the new EMIR data and improve the usefulness of the dataset to authorities in a systematic way.[16]

4.1 Make Refit work (quickly)

As experience with the pre-Refit EMIR data shows, extensively using the data is essential to improving data quality. The authorities should assign sufficient resources to the cleaning and analysis of this information for the exercise of their tasks, even if it constitutes a significant investment of time and effort at the beginning. Only in this way can the systematic shortcomings be identified and addressed. Sharing these identified problems with other authorities greatly accelerates the learning process and leads to a better understanding of the data and its deficiencies. These issues can then be shared with TRs, ESMA, NCAs, or reporting agents, depending on who is best suited to address them.

4.2 Provide aggregated feedback to reporting agents

Trade repositories (TRs) should be mandated to provide a set of simple, key aggregates on the information reported under EMIR to their clients. This measure would enable reporting entities to easily compare the reported information with the data they utilise for their risk management purposes. By doing so, the entities could quickly spot systematic issues affecting their reporting.

4.3 Focus on what is important

The quality issues in EMIR are spread throughout the reporting population, and the sheer size of the dataset makes it close to impossible to deal with all shortcomings at once. The problems, however, frequently follow specific patterns and are often particularly concentrated within a particular set of counterparties and products. The supervisors should start by targeting a few of the largest entities responsible for misreporting or some of the specific instruments where entities struggle to provide high-quality reporting. In this way, noticeable improvement in data quality may be achieved, even with a relatively limited effort.

[16] For more details on these proposed measures, see Agostoni et al. [2024a].

4.4 Provide clear guidelines and foster standardisation

Refit represents a significant advancement with regard to the standardisation of reporting and provision of comprehensive guidance to TRs and market participants. Nevertheless, further steps can be taken to enhance its impact. We advocate that the TRs should be provided with algorithmic, machine-readable guidance that clearly specifies how data should be transformed and aggregated (e.g., for the compilation of the trade and margin reports). This would remove ambiguity in the TRs' processing of EMIR information and ensure that data coming from different TRs can be easily compared and aggregated. Furthermore, making these guidelines public would ensure transparency and allow the community to quickly identify errors or ambiguities.

4.5 Improve reporting of valuations and margins

The data on contract valuations and margins reported under EMIR are critical for comprehending the flow of risk within the financial system. The use of this information, however, presents significant challenges to authorities due to differing practices and limited standardisation. To enhance the accuracy and utility of this information, more detailed guidelines should be established in cooperation between authorities and market participants, taking into account the global context of derivatives trading. It is also worth noting that the consistent reconciliation of margin data between counterparties is of paramount importance.

5 Conclusions

Granular data on derivatives collected under EMIR, despite their shortcomings, have been successfully used for several years to support the tasks of various authorities in the EU. The combined efforts of the regulators have led to an increase in the quality of the reported data and a better understanding of this complex dataset. Still, the quality remains a major barrier for analysts to make full use of this rich and timely source of information.

The introduction of EMIR Refit brings a range of opportunities, but also new challenges. As we have shown, this complex change has led to new issues within the reporting framework, temporarily affecting the capabilities of regulators to use these data for policy work. The value of the daily information, however, is steadily improving, and we believe that it will soon be used to the same extent, if not greater, than before this significant regulatory change.

At the same time, we see further measures that can be applied by regulators, TRs, and reporting agents to improve the quality of the data. These actions do not necessarily involve immense effort among the stakeholders but may address systematic issues affecting the dataset. Better data quality, in turn, will lead to further uptake in the use of these data and better use of automated tools to analyse data at scale.

Bibliography

Agostoni, G., L. Calabrese, M. D'Errico, L. Henkel, and G. Skrzypczynski (2024a): "A small step for EMIR, a giant leap for transparency – first steps towards identifying strategies for boosting data quality in granular data on derivatives," *European Conference on Quality in Official Statistics 2024, forthcoming*, https://airdrive.eventsair.com/eventsairwesteuprod/production-leading-public/d8885e71833e4f01b1f1df9f7aeb4fab.

Agostoni, G., L. De Charsonville, M. D'Errico, C. Leonte, and G. Skrzypczynski (2023): "Anomaly intersection: disentangling data quality and financial stability developments in a scalable way," in *IFC-Bank of Italy Workshop on "Data Science in Central Banking: Applications and tools"*, Bank for International Settlements, no. 59 in IFC Bulletin, https://www.bis.org/ifc/publ/ifcb59_32_rh.pdf.

Agostoni, G., A. Ianiro, A. Jukonis, F. Lenoci, E. Letizia, and G. Skrzypczynski (2024b): "Using trade-level derivatives data for macroprudential analysis," in *64th ISI World Statistics Congress*, https://isi-web.org/sites/default/files/2024-03/ottawa-2023_ips-1077-Grzegorz%20Skrzypczynski.pdf.

ESMA (2022a): "EMIR and SFTR data quality framework 2021," https://www.esma.europa.eu/sites/default/files/library/esma74-47-607_2021_emir_and_sftr_dq_report.pdf.

——— (2022b): "Final Report – Guidelines for reporting under EMIR," https://www.esma.europa.eu/sites/default/files/library/esma74-362-2281_final_report_guidelines_emir_refit.pdf.

——— (2024): "2023 report on quality and use of data," https://www.esma.europa.eu/sites/default/files/2024-04/ESMA12-1209242288-852_2023_Report_on_Quality_and_Use_of_Data.pdf.

Perez-Duarte, S. and G. Skrzypczynski (2019): "Two is company, three's a crowd: automated pairing and matching of two-sided reporting in EMIR derivatives' data," in *IFC conference "Are post-crisis statistical initiatives completed?"*, Bank for International Settlements, no. 49 in IFC Bulletin, https://www.bis.org/ifc/publ/ifcb49_51.pdf.

Assessing initial and variation margins reported under EMIR

Alessandro Pioli[a,1] and Francesco Vacirca[a]

[a] *Directorate General of Market Infrastructure and Payments, ECB, Frankfurt am Main, Germany*

Abstract. In this work, we adopt an internally developed margin simulator to validate initial and variation margins required by central counterparties (CCPs) clearing interest rates swap (IRS) derivatives and reported in the ECB EMIR dataset. The internal Python-based simulator harnesses the power of the ECB's Statistical Data Warehouse and QuantLib functionalities to execute curves bootstrapping, derivatives valuation, and scenario analysis. The validation process aims at computing the contract value and the historical market risk component of classic initial margin methodologies used by the main CCPs, including CCPs clearing IRS.

Key words: EMIR data, margin simulator, interest rates swaps, validation.

1 Introduction

Mandatory clearing in Europe is a key regulatory measure introduced in response to the *great financial crisis*. This requirement, set forth by Article 4 of the European Market Infrastructure Regulation (EMIR), mandates that certain over-the-counter (OTC) derivatives must be cleared through a central counterparty (CCP). The policy aims to enhance transparency and reduce systemic contagion risks in cases of counterparty default. The CCP serves as an intermediary among its participants, referred to as clearing members, effectively becoming the buyer to each seller and the seller to each buyer. To mitigate the risk of counterparty default, CCPs require clearing members to provide collateral in the form of variation margins (VMs), initial margins (IMs), and contributions to a default fund.

[1] This document should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

2 Margin requirements in CCPs

Variation margins (VMs) are collected and distributed to reset current market exposure to zero, thereby accounting for changes in market prices. VMs can assume either positive or negative values and constitute a backward looking element within the clearing framework, as they reflect daily market fluctuations. This *mark-to-market* mechanism is designed to limit the accumulation of losses in derivatives by recalibrating their balances daily. Positive VMs are debited from clearing members with positions incurring losses, while negative VMs are credited to members with profitable positions. VMs are calculated at the portfolio level for derivatives, with the total VM payable or receivable by a clearing member representing the aggregate variation in value across all contracts within that portfolio.

Understanding IM and VM requirements is one of the most effective means of assessing the risk associated with a specific portfolio and is therefore essential for comprehensive risk analysis.

2.1 CCPs margin models

While variation margins are calculated by determining the daily mark-to-market value of each contract within a clearing member's portfolio, initial margins are determined through modelling. Article 41 of EMIR establishes specific requirements that CCPs must follow when setting IM levels, including minimum thresholds for confidence intervals, time horizons, margin periods of risk, and portfolio margining, as well as mandatory inclusion of stress periods. Furthermore, the associated Regulatory Technical Standards (RTS) mandate measures to mitigate procyclicality. Despite these regulations, CCPs retain considerable flexibility to differentiate their margin models by adjusting the numerous parameters involved in the calculations.

2.2 Standard IM model

The methodology for determining IM requirements varies across CCPs but generally consists of three key components:

1. **historical market component**: this component employs traditional risk measures, such as value at risk (VaR) or expected shortfall (ES), calculated over a specified observation period (look-back period);

2. **stress market component**: this element replicates risk calculations across a smaller set of predefined stress scenarios, which may include either synthetic or historical events, such as the Covid-19 market downturn or the European bond crisis;

3. **add-on**: representing the complexities of liquidating certain products in the portfolio under stressed market conditions, this element is usually derived from surveys of market participants and expert assessments.

The historical market component is the foundation of most IM models, often relying on a distribution of historical returns to estimate risk metrics. This nonparametric approach, known as historical simulation (HS), is straightforward to implement and interpret but lacks the ability to model volatility. Volatility modelling, however, is crucial, as near-term returns are significantly influenced by volatility.

A common enhancement to the HS method is the filtered historical simulation (FHS).[2] In FHS, volatility is first estimated for the return series, often using weighting models where recent returns are given greater influence than older ones via a decay parameter. Historical returns are then *de-volatized* and *re-volatized* based on the latest volatility estimates. This approach refines the traditional HS model by incorporating volatility dynamics while maintaining simplicity.

3 Initial and variation margin simulator

In this study, we employ an internally developed initial margin (IM) and variation margin (VM) simulator to validate the IM and VM requirements[3] set by CCPs for clearing interest rate swap (IRS) derivatives, as reported in the ECB EMIR dataset.[4] The internal simulator (see Roulund et al. [2024]), built using Python, leverages the ECB's Statistical Data Warehouse and QuantLib[5] functionalities to perform curve bootstrapping, derivatives valuation, and scenario analysis (see Bianchetti [2009]). For operational efficiency and simplicity, valuations are conducted on a self-discounted basis, thereby avoiding the complexities associated with a multi-curve framework. This simulator facilitates the investigation and application of multiple use cases, including:

- anti-procyclicality tools testing;
- margin model back-testing;
- data validation;
- stress testing.

3.1 Replicate FHS with the simulator

The focus of the work is on OTC interest rate swaps (IRS). To replicate FHS, the simulator generates the zero-coupon yield curve (zero-curve) for each day of the look-back period and computes market returns over a pre-defined period, the market period of risk (MPOR), usually 5 days for OTC traded derivatives. Once the zero curve is created for each day, the volatility is modelled as an exponential weighted moving average (EWMA) process. The historical returns are then "filtered" with the current volatility. This is achieved by removing the original volatility (de-volatize) of each return scenario and applying the current volatility

2 See Gurrola-Perez and Murphy [2015].
3 VM validation is implicitly done by validating the reported contract value of the transaction.
4 The ECB EMIR database includes information on euro area derivative contracts for which at least one counterparty is based in the euro area.
5 See https://www.quantlib.org/.

(re-volatize) of the scenario based on the latest day. Once scenarios are generated, the IM are calculated by taking the value at risk or expected shortfall of the portfolio given the scenarios within the predefined confidence interval.

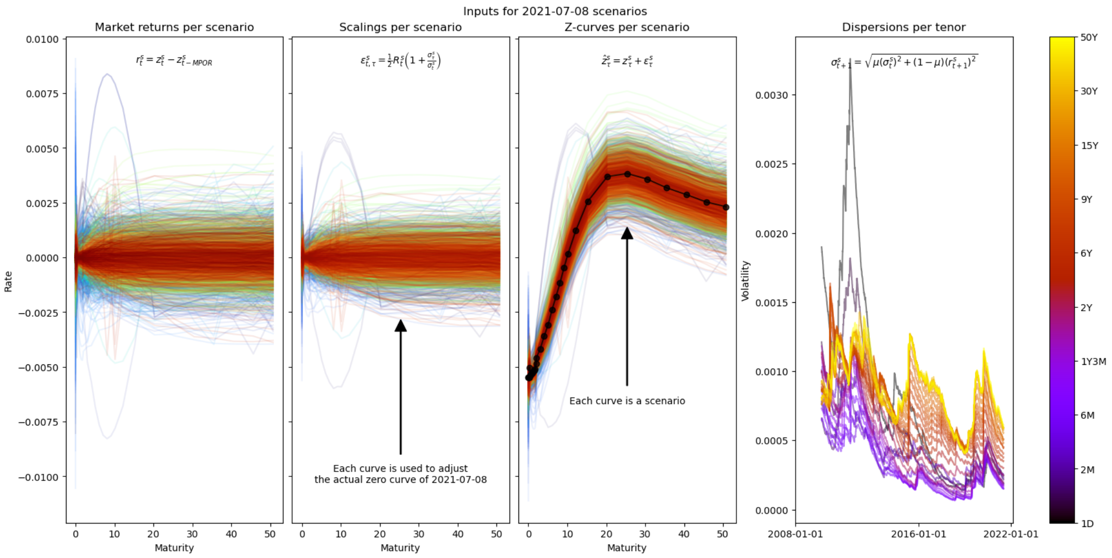


Figure 1 – FHS scenario generations on 3500 days lookback period starting on the 2021-07-08.

The simulator is highly customizable, allowing adaptation to different CCP methodologies by adjusting the following parameters:

- tenors: points used for the bootstrapping of the yield-curves;
- λ: decay parameter for the EMWA;
- look-back period: number of historical dates per scenario;
- MPOR (minimum set by EMIR);
- filtering: enable, or not, historical returns filtering;
- confidence level: for the risk metrics (minimum set by EMIR);
- scaling parameter: value used to scale up the risk metric to a higher confidence interval;
- risk metric: value at risk or expected shortfall.

A graphical example of a scenario generation process is pictured in Figure 1.

4 Analysis

4.1 Goal of the analysis

This study not only contributes to the validation of data reported under EMIR but also advances the development of an IM simulator, providing valuable insights into the global discourse on IM transparency and the role of margin simulators in enhancing liquidity preparedness for participants in the derivatives markets. The validation process focuses on calculating contract valuations and the historical market risk component, a core element of traditional IM methodologies employed by leading CCPs for OTC interest rate swaps.

4.2 Caveats

For the sake of efficiency and simplicity, the simulator has some known limitations that include: (a) derivatives valuation conducted on a self-discounted basis: as previously mentioned this trades simplicity for accuracy; (b) ability to price only euro denominated derivatives: due to the richness of available data on the instrument needed to bootstrap the yield curves; (c) lack of add-ons estimation: as this is usually derived using a mix of expert judgment and market participant surveys.

A full validation of initial margins reported under EMIR would require replication of all model components, including add-ons, which falls beyond the scope of this analysis. As this study concentrates on the primary component of margin (i.e., the historical market risk element) some variance from the reported values is anticipated. Our estimates may be considered a lower bound, as they do not encompass add-ons or additional stress periods, both of which are additive elements in the overall margin calculation.

EMIR fields used by the simulator
valuation_timestamp
reporting_cpty_id
contract_value_currency
counterparty_side
contract_type
asset_class
notional
fxd_rate_leg1
fxd_rate_leg1_day_count
notional_currency1
effective_date
maturity_date
fxd_rate_leg1_pymnt_frq_mltplr
fxd_rate_leg1_pymnt_frq_prd
flt_rate_leg2
flt_rate_leg2_pymnt_frq_mltplr
flt_rate_leg2_pymnt_frq_prd
flt_rate_leg2_benchmark_value
contract_value_currency
collateral_portfolio_code

Statistics on analysed portfolios	
Number of portfolios	305
Average number of trades per portfolio	25
Average portfolio's notional	1.42 Bn €
Number of portfolios containing compressed trades	49
Average remaining maturity	17.6 Years
Average fix rate	1.65%
Euribor benchmark rate at time of analysis (8/1/24)	3.929%

Figure 2 – EMIR fields used by the simulator (left) and general statistics (right).

5 Dataset

The dataset is composed of 305 portfolios containing only plain vanilla Euribor interest rate swaps. Only a specific subset of information is needed among the more than 300 fields of the dataset (see Figure 2).

6 Results

6.1 Initial margin

Before delving into the results, it is necessary to define a metric to assess the goodness of the validation. For IM we defined the IM ratio as the ratio between the IM estimated via the simulator and the IM reported in EMIR.

Figure 3 – Initial margin ratio general statistics: distribution (top-left), scatterplot over portfolio size (top-right), scatterplot over remaining maturity split by compression (bottom).

The distribution is concentrated around the 88.3% average value, the interquartile range is contained between 83.8% and 98.2%. The results are in line with expectations; as mentioned above, we are estimating a single component of the margin model and therefore under-estimating by design. Despite the promising results, further investigation is needed to understand the outliers. The size of the portfolio and compression do not seem to impact the results while there tend to be better results for portfolios with remaining maturities around 25 years; this could be explained by the choice of financial instruments used to bootstrap the short end of the zero curve.

6.2 Variation margin

As VMs represent the flow of cash that is exchanged when each contract in the portfolio is marked-to-market, validating them would have required following up the evolution of the portfolios over multiple days. To simplify the endeavour, we decided to use the contract value as a proxy. We defined the net present value

(NPV) ratio as the ratio between the sum of the contract values reported in EMIR (within a portfolio) and their estimated NPV via the simulator.

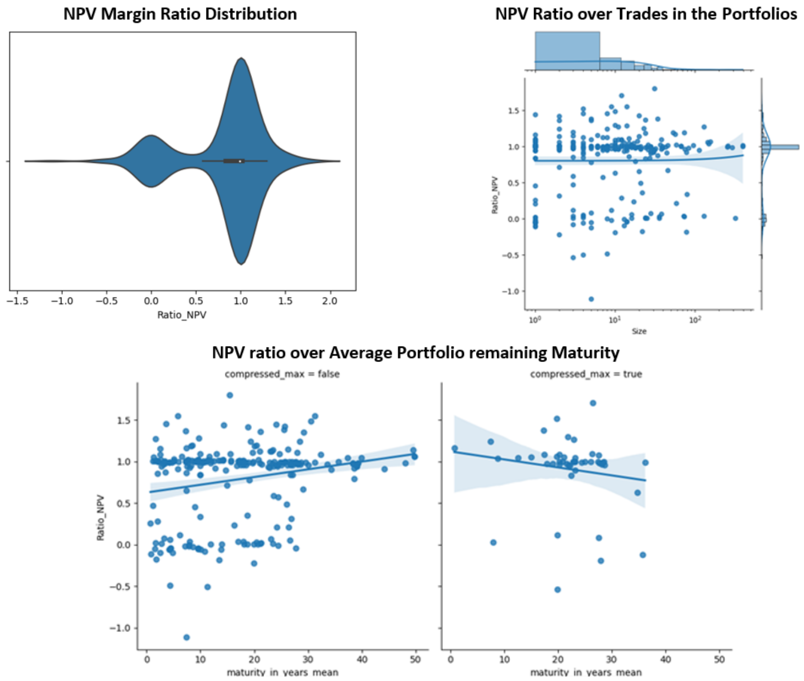


Figure 4 – NPV ratio general statistics: distribution (top-left), scatterplot over portfolio size (top-right), scatterplot over remaining maturity split by compression (bottom).

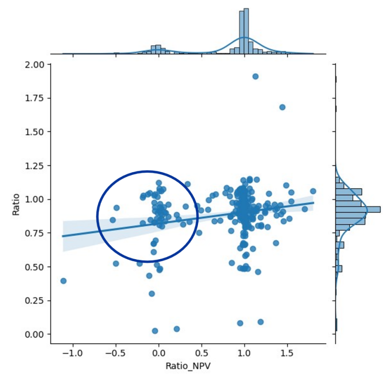


Figure 5 – Initial margin ratio over NPV Ratio.

The distribution is bimodal with two peaks around 0% and 100%, and the interquartile range is between 83.3% and 101.9%. The peak around 0 is given by the fact that, for a significant number of portfolios, the simulator NPV is far greater than the reported contract value. Remaining maturity, compression and size of the portfolio do not impact results. To further investigate this result, we check the value of both ratios for the same portfolios (see Figure 5).

It is interesting to note that most of the portfolios for which the NPV Ratio lays around 0, i.e. the dots that lie within the blue circle in the Figure 5, have an IM Ratio that falls in what we can consider a good range (between 75% and 125%).

7 Conclusions

Further investigations are required to identify the root causes of the observed outliers. The elevated validation metrics for the IM may be attributed to the fact that the values are reported at a portfolio level and are consequently less influenced by the accurate reporting of all transaction parameters.

In conclusion, the results of the analysis of the initial 305 portfolios are highly promising, both in terms of validating the simulator and the dataset.

Additional analyses utilizing the simulator are currently underway, with a particular focus on testing anti-procyclicality (APC) tools.

Bibliography

BIANCHETTI, M. (2009): "Two curves, one price: Pricing & hedging interest rate derivatives decoupling forwarding and discounting yield curves," *arXiv*, https://arxiv.org/abs/0905.2770.

GURROLA-PEREZ, P. AND D. MURPHY (2015): "Filtered historical simulation value-at-risk models and their competitors," in *Staff Working Paper*, Bank of England, No 525, https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2015/filtered-historical-simulation-value-at-risk-models-and-their-competitors.pdf.

ROULUND, R., F. VACIRCA, AND R. KOCI (2024): "An objective approach to limit initial margin procyclicality," in preparation.

How fair is the value of contract reported under EMIR? An analysis on interest rate derivatives

Michele Leonardo Bianchi[a,1] and Dario Ruzzi[a]

[a]*Financial Stability Directorate, Banca d'Italia, Roma, Italy*

Abstract. We conduct an extensive data quality analysis of the "value of contract" field (i.e., the fair value) reported in EMIR for a large sample of interest rate derivative positions of Italian banks. We use granular regulatory data on euro interest rate swap trades between January 2021 and June 2023 to assess the quality of this field. Drawing on our swap pricing model, we document that despite model inaccuracy in pricing a not negligible number of trades, the errors in fitting swap value changes over time are limited. This empirical finding enables authorities to perform comprehensive sensitivity analyses and stress tests based on the detailed and granular data provided under EMIR.

Key words: EMIR, interest rate derivatives, data quality, fair value.

[1] The views in this paper are the authors' and do not necessarily reflect those of the Banca d'Italia. Errors and omissions are the authors own responsibility.

1 Introduction

To date, there have been relatively few empirical investigations of granular derivatives exposures, primarily due to limited data availability (see, the ESMA [2023] Annex). The implementation of the European Market Infrastructure Regulation (EMIR) in 2014 requires EU entities to report on a daily basis their derivative transactions to Trade Repositories (TRs), resulting in improved data availability. TR data are regulatory (non-public) contract-level data; for instance, in Europe, EMIR data are accessible only to the European Securities and Markets Authority (ESMA), European Systemic Risk Board (ESRB), European Central Bank (ECB), and national competent authorities. Furthermore, they come in high volume and therefore a big data infrastructure is required for their storage, manipulation, and analysis. The combination of confidentiality requirements and technical complexities has slowed down the development of a specific research avenue that leverages the granular derivatives exposures of market participants. Notably, scholarly articles investigating EMIR and similar regulatory data have only recently emerged in peer-reviewed academic journals.

Research focusing on granular derivatives exposures can be categorised into four main areas: (1) data description, encompassing data quality concerns; (2) deep dive into specific contract types, counterparties, or market occurrences; (3) systemic risk evaluation of margin practices and network structures; and (4) risk sensitivity analysis. Our work complements the first strand of literature with a view toward the fourth.

The quality of data reported under EMIR is critical for effectively monitoring financial stability risks. Persistent data quality issues can significantly limit the use of these data, undermining the ability to conduct analyses and provide reliable insights (see ESRB [2022] and ESMA [2023]). Ensuring high-quality input data is therefore essential for achieving accurate and meaningful outcomes.

Motivated by this fact, in this work we propose a framework to assess the quality of data reported under EMIR. We examine the accuracy, consistency, and reliability of the reported interest rate swap values by performing full revaluation (i.e. pricing) of the swap contracts and by comparing model-implied swap values to those made available in reporting to authorities. More in detail, we perform a comprehensive data quality assessment of the "value of contract" field reported under EMIR for a broad sample of interest rate derivative positions held by Italian banks. Using data on euro interest rate swap transactions from January 2021 to June 2023, we evaluate the accuracy and reliability of this field and we assess how fair it is relative to our pricing model.

Our data quality analysis of the granular derivatives information available to regulators unveils that it is easier to reconcile time changes in the reported swap values, rather than the values themselves, with the contract specific characteristics that are also being reported. This finding validates the computation of entity's interest rate risk exposure in swaps (i.e. swap sensitivity) as the change in value of its swap portfolios after a shock in the yield curve (see Bianchi et al. [2024]). This can be done by repricing each individual swap contract before and after a shock, and then taking the difference between the pre- and post-shock values. This is a full revaluation approach that delivers changes in market values of contracts computed using comprehensive pricing formulas.

The remainder of the paper is organised as follows. Section 2 describes the EMIR data analysed in this study. In Section 3 we show how we compute swap fair values and we report the main empirical findings on EMIR data quality checks. Section 4 concludes.

2 EMIR data

This paper employs transaction-level EMIR data to analyse interest rate derivative trades reported by Italian banks. EMIR data have been accessible to Banca d'Italia as of December 29, 2020, covering entities and markets within its jurisdiction and financial stability mandate.[2] With over 100 data fields, the dataset provides detailed information on each transaction, including counterparty identity, derivative type, contract value, maturity, outstanding notional amount, underlying, execution and clearing venues, and any collateral (margin) exchanged.

We focus on a subsample of euro-denominated interest rate derivatives for Italian banks consolidated at the group level.[3] Although EMIR data are recorded daily, with approximately 30 million records per trading day, for practical reasons we conduct our analysis at a weekly frequency. Specifically, we sample trades outstanding on the Wednesday of each week from January 2021 to June 2023, using "trade state" data, which includes all pending trades at the end of a given day.

EMIR reporting is dual-sided, meaning both counterparties to a derivative contract are required to report if they fall under EMIR's scope. As a result, our dataset may contain duplicate reports for the same trade, submitted by each counterparty. To address this, we use the "trade id" field to identify duplicates and retain only one report per trade to avoid double-counting. For each entity of interest (e.g., a bank), when two reports are available, we prioritise the one submitted by that entity.

We now detail the selection of contracts analysed in the empirical section. Our focus is on forward rate agreements (FRA) and interest rate swaps (IRS) referenced to Euribor, and overnight index swaps (OIS) referencing Eonia and €STR.[4] Among these, Euribor swaps are the most traded and liquid derivatives for hedging euro-denominated interest rate risk (see Grassi et al. [2022]). We analyse single-currency, fixed-for-floating contracts, including both spot-starting and forward-starting trades.[5] For simplicity, we use the term "swaps" to collectively refer to IRS, OIS, and FRA, as FRA are essentially single-period IRS with a forward start date.

---

2 Bianchi et al. [2025] provides a detailed description of the EMIR dataset, the portion accessible to Banca d'Italia, and the framework used to process raw data into a clean format suitable for analysis, addressing data quality and practical challenges.

3 Consolidation relies on Banca d'Italia's regulatory group structure for Italy-based parent banks and on GLEIF (Global Legal Entity Identifier Foundation) data for other groups.

4 Eonia was discontinued on January 3, 2022, but OIS contracts referencing it remain active until expiration. We treat these as equivalent to €STR-based OIS.

5 In forward-starting swaps, payment exchanges begin at a future date T_1 and continue until maturity T_2. A long (pay fixed-rate) position in such a swap is equivalent to holding a spot-starting swap maturing at T_2 and shorting one maturing at T_2.

We select in EMIR the trades where:

- the "asset class" field is equal to "IR" (for "interest rate");
- the "contract type" field is equal to "SW" (for "swap") or "FR" (for "forward rate agreement");
- either the "floating rate of leg 1" field or the "floating rate of leg 2" field is not blank, and contains the term "euri", "str", "eona", or "Eonia" (after lowercasing the text);
- either the "fixed-rate of leg 1" or the "fixed-rate of leg 2" field is not blank;
- only one of the "notional currency 1" and "notional currency 2" fields is not blank, or they both contain the same value;
- the "effective date" is after the "reference date"(i.e., forward-starting contracts);

and we collect the contractual features of each contract.

In a fixed-for-floating swap, one party pays a fixed rate and receives a floating rate on a notional amount for a fixed term, while the other party does the opposite. We adopt the following convention: the counterparty paying the fixed rate is the swap buyer (holding a "long" position), while the counterparty receiving the fixed rate is the seller (holding a "short" position). We determine long and short positions using the "counterparty side" field, which is marked as "B" for buy or "S" for sell. Regulation EU/2013/148 defines the buy side as the payer of leg 1 and the sell side as the payer of leg 2, where each leg corresponds to either the fixed or floating rate. Regulation EU/2017/105 later establishes that in interest rate or inflation swaps, the fixed-rate payer is the buyer, and the fixed-rate receiver is the seller. To account for this regulatory change, we identify the reporting counterparty's role by repricing the contract under both interpretations. We follow the more recent regulation unless this interpretation results in a contract value with a sign opposite to that reported in EMIR. In practice, over 95% of sampled trades align with the latest regulation.

To ensure consistency, we filter Euribor-referencing trades to include only contracts where the floating rate's maturity is 1, 3, 6, or 12 months and this is equal to the number of months between two consecutive floating rate payment dates.

For the fixed rate, our data cleaning process begins with the "price/rate" field, which, when valid, provides the swap rate (i.e., the fixed rate) as a percent value. If unavailable, we then use the "fixed-rate of leg 1" and "fixed-rate of leg 2" fields. We compare the reported fixed rate of each contract with the prevailing par swap rate (sourced from LSEG) on the contract's effective date for a term matching its original maturity. This comparison helps to identify cases where the reported rate is not expressed as a percentage as well as to detect outliers (below -2.5% or above 25%), which we replace with the market par swap rate.

After applying these data filters and excluding trades with valuations dated more than a week before the reporting date, our final sample comprises 218,717 unique swap contracts traded by 54 banks, totaling 14,664,747 swap-week observations from January 2021 to June 2023.

3 Empirical findings

In this section we present our empirical results. In Appendix A.1 we provide detailed information on how to bootstrap the riskless spot (i.e. zero-coupon) curve that provides discount rates for swap pricing. The comprehensive pricing formulae of swaps are provided in Appendix A.2 and are based on the interpretation of a contract as a long/short combination of a bond paying the fixed-rate on the swap and a floating-rate bond paying the money market reference rate. This means that a short (i.e. receive-fixed) swap position can be valued as the price of the fixed-rate bond minus the price of the floating-rate bond, that is

$$V^{swap} = B^{fix} - B^{fl} \ . \tag{1}$$

We investigate in Section 3.1 the accuracy of the swap pricing model (1) relative to the values reported in EMIR. We consider the pricing error not only in the contract values but also in their time-series changes, as a small error in the latter is what matters most in sensitivity analysis (see Bianchi et al. [2024]).

3.1 Swap pricing fit

We start the assessment of the swap pricing model (1) by comparing the model-implied contract values, which we denote by $\widehat{CV}$, to the values reported in EMIR, which we denote by CV. The goodness of the model in fitting the swap contract values is shown in Panel (a) of Figure 1, where each dot represents a swap-week observation in our dataset.

Overall we find a good fit between the data and the model, as indicated by the almost unitary slope (1.001) of the estimated line obtained with robust regression (the M-estimation method). At the same time, we observe a relatively large amount of dots departing from the regression line due to substantial pricing errors (we highlight in green the observations with absolute pricing errors greater or equal to €25 million). Nevertheless, it is reassuring that these dots are mostly located in the first and third quadrants, thus indicating a good correspondence of sign between the model-implied and reported contract values. It is important to note that pricing errors can be the result not only of erroneous inputs fed into the model (e.g. a fixed-rate measured on the wrong scale or an unreported spread paid on the floating leg) but also of erroneously reported contract values against which we compare the model estimates. Furthermore, some discrepancies are possibly due to adjustments made to the fair value of derivatives contracts that the trading entities include in their pricing framework to take into account funding, credit risk and regulatory capital costs. These potential trade-by-trade valuation adjustments are not accessible to us.

To help visualise the magnitude of the pricing errors, we show the time trend of the interquartile range – that is the 25th, 50th, and 75th percentiles of the distribution – of their absolute values in Panel (b) of Figure 2. Calculated across all dates, the median absolute error is €21,182. At the end of the sample, the value is €26,130, considerably higher than the €17,438 observed at the start. The increase is at least in part a result of the larger contract values reported in EMIR since 2022, when interest rates began to rise steadily. In fact, as can be seen from Panel (a) of Figure 2, the median absolute contract value in EMIR has almost doubled since the start of year 2021 to €225,737.

(a) Contract values (b) Weekly changes in contract values

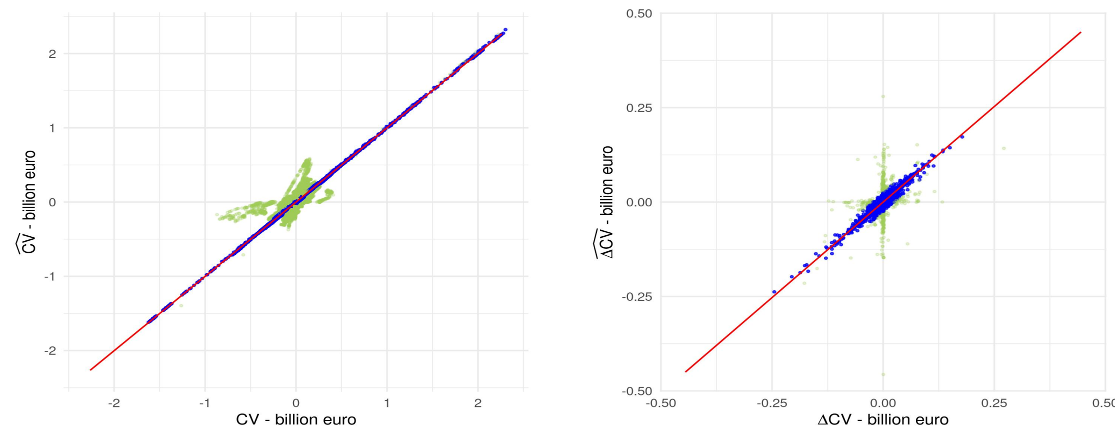


Figure 1 – Goodness of fit of swap pricing model. The figure displays in Panel (a) the fit between the swap contract values (CV) reported in EMIR, and those implied by the pricing model (1), $\widehat{CV}$, and in Panel (b) the weekly changes in the swap contract values (ΔCV) reported in EMIR, and those implied by the pricing model (1), $\widehat{\Delta CV}$. In Panel (a) each dot represents a swap-week observation in the sample (14,664,747 in total). Green dots denote observations (43,867 in total) with absolute pricing error greater or equal to €25 million. The regression line, in red, is estimated to have slope of 1.001 with robust method. In Panel (b) each dot represents a swap-week observation (14,441,405 in total). Green dots denote observations (458 in total) with absolute pricing error greater or equal to €25 million. The regression line, in red, is estimated to have slope of 1.012 with robust method.

Since sensitivity analysis can provide indications about asset price changes, we now discuss the accuracy of model (1) in estimating the change in value of a swap position between two consecutive dates. Let ΔCV and $\widehat{\Delta CV}$ denote the weekly changes in, respectively, the EMIR reported and model-implied swap contract values. The goodness of the model in fitting the weekly value changes is shown in Panel (b) of Figure 1, where each dot represents a swap-week observation in our dataset.

Overall we find that the model fit is improved when the change in value of a swap position, rather than the value itself, is considered. Graphically this means that the vast majority of points in Panel (b) of Figure 1 lie on, or are visually indistinguishable from, the regression line, which is estimated to have a slope of 1.012.[6] As indicated by the less pronounced presence of green dots signalling outliers, the model fitting errors appear smaller than in Panel (a) of Figure 1; this fact is also confirmed by Panel (b) of Figure 3, which displays the time trend of the interquartile range of the absolute errors in estimating swap value changes. Over the full sample period, the median absolute error is €4,330, while the 25th and 75th percentiles amount to €706 and 25,266, respectively. The volatility over

[6] A quick inspection of the outliers reveals that misreporting in EMIR is a contributory factor of the largest deviations observed.

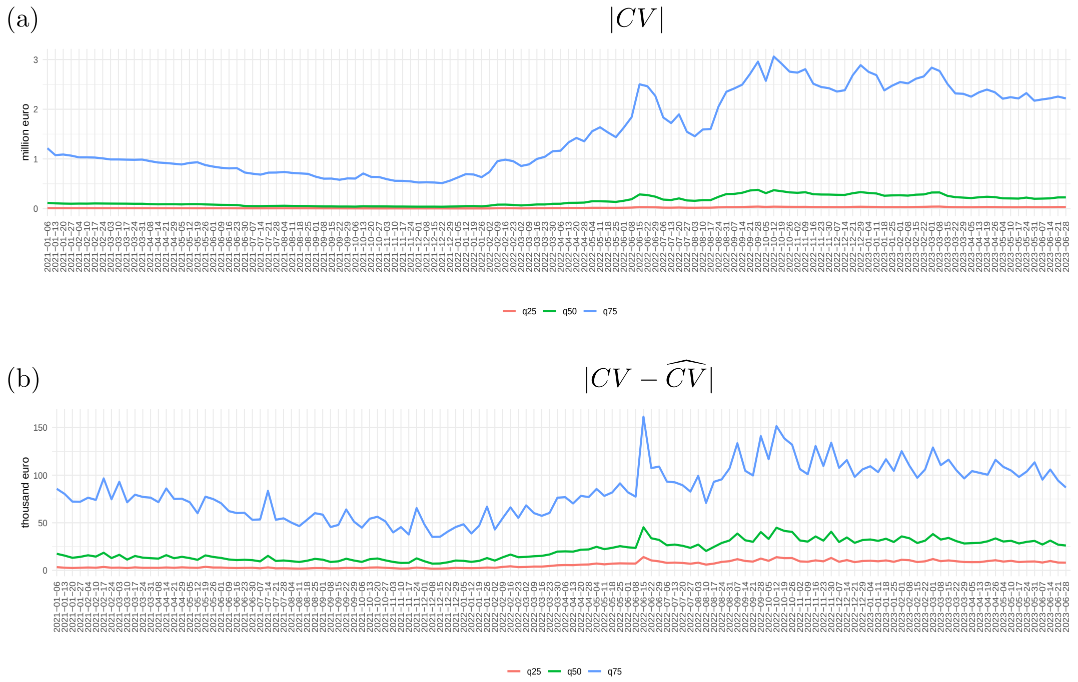


Figure 2 – Swap contract values and model fitting errors. Panel (a) displays the time trend of the 25th, 50th, and 75th percentiles of the distribution of the absolute contract values reported in EMIR. Panel (b) displays the time trend of the percentiles of the absolute difference between the contract values reported in EMIR and those implied by the model (1).

time in the interquartile range of the errors seems to track closely the volatility in the interquartile range of the weekly changes in the EMIR swap values, which is shown in Panel (a) of Figure 3.

So far we have analysed the performance of the swap pricing model (1) at the level of individual contracts. We now adopt a broader perspective to investigate the model fit on the aggregate swap portfolio value of the Italian banking system, that is the sum of the values of all swap positions of the Italian banks in the sample on a given date. The time trends of the observed and model-implied values of the aggregate swap portfolio are displayed in Panel (a) of Figure 4. One can clearly see that not far from the start of the sample a spread of about €2.5 billion appears between the two series and remains relatively stable throughout the sample. These systematic deviations point to the need for further investigation of the derivatives information collected in EMIR, potentially with the involvement of the reporting entities. On the positive side we show, in Panel (b) of Figure 4, that the pricing fit is extremely satisfactory when the weekly changes of the swap portfolio values are considered. The observed and model-implied series track each other quite closely and are often visually indistinguishable.

Taken together, the results of this section suggest that, despite the model's inaccuracy in pricing a not negligible number of trades, the errors in fitting swap value changes over time are significantly reduced compared to those in fitting the values themselves. This provides compelling evidence that validates sensitivity analyses based on model-implied changes in swap values following an interest rate shock.

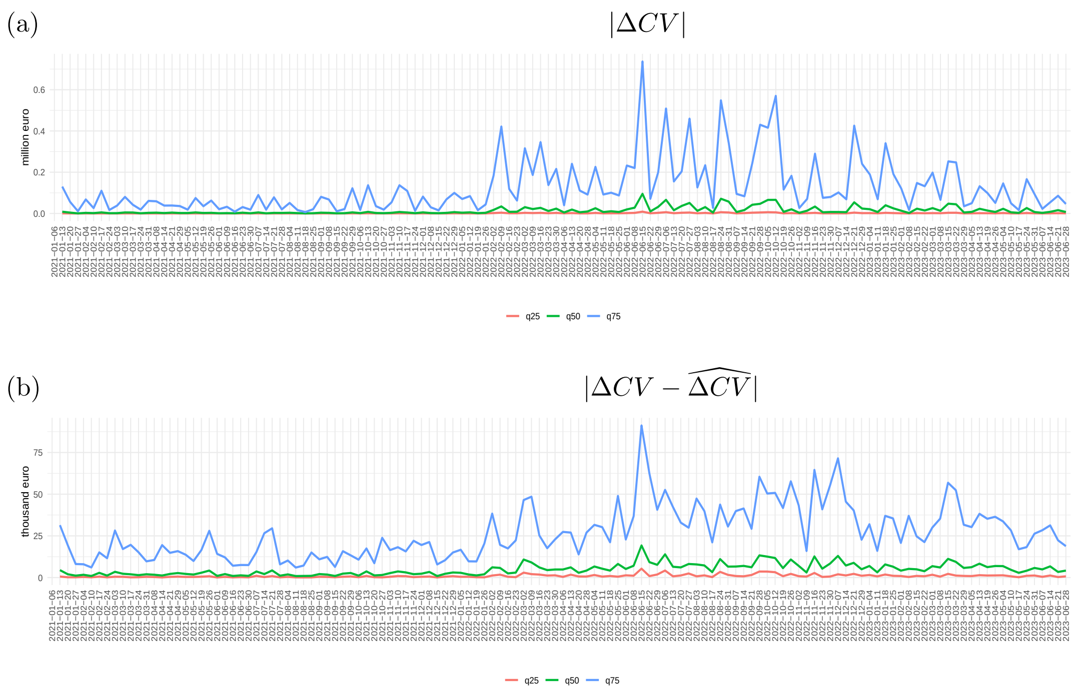


Figure 3 – Weekly changes in swap contract values and model fitting errors. Panel (a) displays the time trend of the 25th, 50th, and 75th percentiles of the distribution of the absolute value of weekly changes in contract values reported in EMIR. Panel (b) displays the time trend of the 25th, 50th, and 75th percentiles of the absolute difference between the weekly changes in the contract values reported in EMIR and those implied by the model (1).

(a) Swap portfolio values

(b) Weekly changes in portfolio values

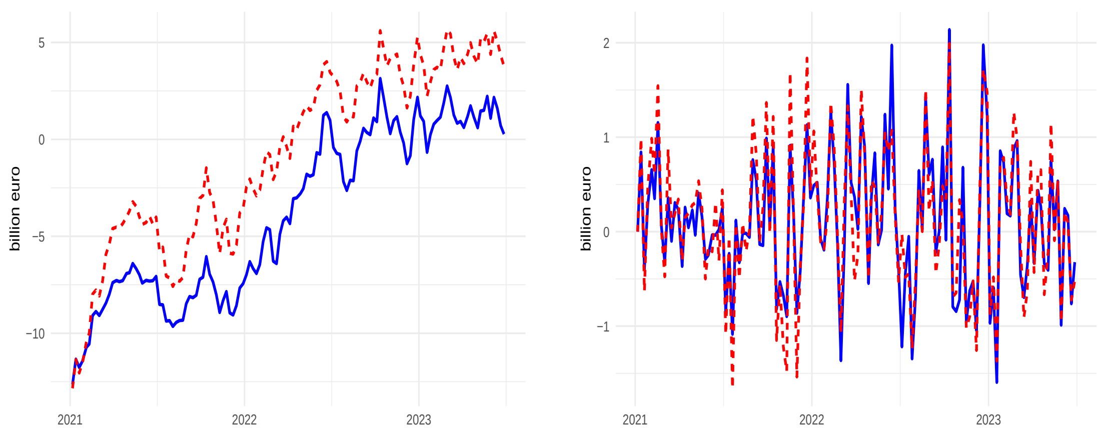


Figure 4 – Swap portfolio values of the Italian banking system and model fitting errors. Panel (a) displays the values of the Italian banks' aggregate swap portfolio observed in EMIR (solid blue line) and implied by the model (1) (dashed red line). Weekly changes in the aggregate swap portfolio values observed in EMIR (solid blue line) and implied by the model (dashed red line) are reported in Panel (b).

4 Conclusions

In this paper we develop a tool to assess the fairness of the “value of contract” of derivatives positions reported in EMIR. Our framework relies on granular information on interest rate swaps contracts, including contract-specific characteristics. The investigation of the data quality issues associated with this field has revealed a critical area for improvement. Our contract-by-contract full revaluation approach allows us to conclude that, despite model inaccuracy in the pricing of a not negligible number of trades, the errors in fitting swap value changes over time are significantly reduced compared to those in fitting the values themselves. Authorities can leverage the detailed and granular data provided under EMIR to perform comprehensive sensitivity analyses and stress tests, enabled by this latter empirical finding.

This work represents a pioneering effort as it is the first extensive data quality analysis conducted on the “value of contract” field. This groundbreaking initiative establishes a foundational benchmark for rigorously evaluating and improving data accuracy in this pivotal aspect of EMIR reporting. Accurate and reliable data are the core of effective risk assessment. Resolving these data quality issues is crucial for sound regulatory oversight and financial stability monitoring.

Bibliography

Bianchi, M., D. Ruzzi, and A. Segura (2024): "Shifting the yield curve for fixed-income and derivatives portfolios," *arXiv*, https://arxiv.org/abs/2412.15986.

Bianchi, M. L., B. Sorvillo, D. Ruzzi, F. Apicella, L. Abate, and L. Del Vecchio (2025): "EMIR data for financial stability analysis and research," in *IFC-Banca d'Italia Workshop on "Data science in central banking: enhancing the access to and sharing of data"*, Bank for International Settlements, no. 64 in IFC Bulletin, https://www.bis.org/ifc/publ/ifcb64_12.pdf.

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.

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.

Grassi, A., T. Kockerols, F. Lenoci, and C. Pancaro (2022): "Euro area interest rate swaps market and risk-sharing across sectors," in *ECB Financial Stability Review*, November, https://www.ecb.europa.eu/press/financial-stability-publications/fsr/focus/2022/html/ecb.fsrbox202211_03~b521c85b4b.en.html.

Hull, J. C. and A. White (2015): "OIS discounting, interest rate derivatives, and the modeling of stochastic interest rate spreads," *Journal of Investment Management*, 13, 64–83.

Smith, D. (2013): "Valuing interest rate swaps using OIS discounting," *Journal of Derivatives*, 20, 49–59.

Appendix

A.1 Bootstrapping the yield curve

We review the procedure for bootstrapping the riskless spot (i.e. zero-coupon) curve that provides discount rates of swap and bond pricing. Following the standard multi-curve method outlined by Hull and White [2015] and Smith [2013], among others, we bootstrap spot rates from €STR-referencing OIS rates, and from Euribor swap rates. To this end, we obtain from LSEG daily rate observations for €STR OIS with maturity up to 30 years, and for 6-month Euribor swaps with maturity up to 50 years. These are "par" (at-market) swap rates representing the fixed-rate paid on bonds valued at par (i.e. price of 100).[7] We linearly interpolate the swap rates to have data for maturities evenly spaced by 3-month intervals as we assume quarterly settlements of the swap contracts and, accordingly, four payments a year on the par bonds. Swaps of up to one-year's maturity have only a single payment at the contract end date, therefore, we use their rates as the spot rates of the corresponding maturity.

To infer the rest of the spot curve that is consistent with the sequence of par swaps we rely on the following bootstrapping technique. Let $z^{(T)}$ be the annualised spot interest rate observed today for maturity T (in months), $s^{(T)}$ be the annualised par swap rate for maturity T, and $P^{(T)}$ be the T-maturity par bond's price, which is equal to its notional amount, $N = 100$. Under the assumption of quarterly settlement, each coupon payment of the par bond amounts to $C = s^{(T)}N/4$. Using the spot rates already available for all maturities before T, the next-in-line spot rate $z^{(T)}$ is the solution to the following bond pricing equation

$$P^{(T)} = \sum_{n \in D} \frac{C}{(1 + z^{(n)} \frac{n}{12})} + \frac{C + N}{(1 + z^{(T)} \frac{T}{12})} \ , \tag{1}$$

where $D = \{3, 6, ..., T-3\}$. Equation (1) implies that the spot rate gets computed as

$$z^{(T)} = \frac{12}{T} \left[\frac{C + N}{P^{(T)} - \sum_{n \in D} \frac{C}{(1+z^{(n)} \frac{n}{12})}} - 1 \right] . \tag{2}$$

We repeat the calculation sequentially until the longest maturity of the par swap rates, and then we linearly interpolate the spot rates to have data for maturities evenly spaced by 1-month intervals.[8] We use the above procedure to bootstrap one spot curve from OIS rates and one from Euribor swap rates. The former spot curve provides discount factors for maturities up to 30 years, which is the longest maturity observed in the €STR OIS market, and determines the forward rates affecting the cash flows of the OIS contracts sampled from EMIR.[9] The latter spot

7 The par swap rates are also the rates that make the market value of the swap contracts equal zero.

8 When discounting the future cash flows of the swap contracts sampled from EMIR, despite using an actual/365 convention, we will round to the nearest month the maturity of the relevant spot rates.

9 Eonia was discontinued on 3 January 2022 and therefore we use €STR rates to value also the Eonia-referencing OIS contracts that continue to exist until their expiration.

curve provides discount factors for maturities longer than 30 years and determines the forward rates affecting the cash flows of the FRA and IRS contracts sampled from EMIR. We compute Euribor and €STR forward rates, which represent the expected future rates on the floating leg of the swap contracts considered in our work, starting from the corresponding spot curve and assuming the absence of arbitrage.

Letting $f^{(T_{i-1},T_i)}$ be the forward rate between times T_{i-1} and T_i (in months) – i.e. the interest rate expected today on a zero-coupon investment starting at time T_{i-1} and ending at T_i – the assumption of no-arbitrage implies that the following equality holds

$$f^{(T_{i-1},T_i)} = \frac{12}{T_i - T_{i-1}} \left(\frac{1 + z^{(T_i)} \frac{T_i}{12}}{1 + z^{(T_{i-1})} \frac{T_{i-1}}{12}} - 1 \right) . \tag{3}$$

A.2 Swap pricing formulae

Besides the fixed and floating rates discussed in Section 2, the other inputs of our swap pricing model (1) consist of the contractual features that are reported in the following EMIR fields: "valuation timestamp", "counterparty side", "notional", "maturity date", "fixed-rate payment frequency", and "floating rate payment frequency". We also pull the "value of contract" field which we use to assess the fit of our pricing functions. As a first step in determining the two bond prices in (1), we define the payment schedule of the fixed and floating legs by using the payment frequency expressed in months and working backwards from maturity date T to valuation date t_0. For each payment schedule, we compute, both in days and in months, the time frame between t_0 and each of the future payment dates.[10] We will use the time frames expressed in days to perform the actual/365 day-count convention for rates when discounting, and those expressed in months to select the maturity of the relevant spot rate from those interpolated at 1-month intervals. We cap all time frames between payment dates and t_0 to 50 years as this corresponds to the longest maturity of the bootstrapped spot rates.

To calculate the value of a swap, we implement the following pricing formulae. Let us consider an interest rate swap referencing Euribor with maturity T and notional amount N. The fixed leg, which pays the annualised swap rate $s^{(T)}$, makes q^{fix} payment(s) per year, for a total of I payments between t_0 and T. Let n_i and d_i be the time frames in months and in days, respectively, between t_0 and the i-th fixed-rate payment date, with $i = 1, 2, ..., I$. The floating leg, which pays the annualised k-month Euribor rate, with $k = 1, 3, 6, 12$, makes q^{fl} payment(s) per year, for a total of J payments between t_0 and T. Let n_j and d_j be the time frames in months and in days, respectively, between t_0 and the j-th floating-rate payment date, with $j = 1, 2, ..., J$. We assume that at each payment date j the floating rate bond pays the reference rate that prevailed on the market at the previous payment date, $j - 1$. This means that the first floating rate payment is already known at valuation date and is based on the Euribor spot rate observed $n_1 - k$ months before

[10] In order to have integer numbers for the number of months in the time frame, we shift forward both the valuation and the next payment dates to month-end dates and we count the months between the shifted dates. If the first next payment date is within a month, we set the time frame to 1 month.

t_0. Using the spot rates $z^{(\cdot)}$ defined in Appendix A.1, the bond price of the fixed leg gets computed as

$$B^{fix} = \sum_{i=1}^{I} \frac{C^{fix}}{(1 + z^{(n_i)} \frac{d_i}{365})} + \frac{N}{(1 + z^{(n_I)} \frac{d_I}{365})} \ , \tag{4}$$

where $C^{fix} = s^{(T)} N / q^{fix}$. At the same time, while relying on both spot and forward rates, the bond price of the floating leg gets computed as

$$B^{fl} = \sum_{j=1}^{J} \frac{C_j^{fl}}{(1 + z^{(n_j)} \frac{d_j}{365})} + \frac{N}{(1 + z^{(n_J)} \frac{d_J}{365})} \ , \tag{5}$$

where for $j = 1$ we use $C_1^{fl} = z_{t_0^-}^{(k)} N / q^{fl}$, with $z_{t_0^-}^{(k)}$ denoting the k-month Euribor spot rate observed $n_1 - k$ months before t_0, and for $j > 1$ we use $C_j^{fl} = f^{(n_{j-1}, n_j)} N / q^{fl}$, with $f^{(n_{j-1}, n_j)}$ denoting the k-month Euribor forward rate between times n_{j-1} and n_j. By combining the two bond prices as per equation (1), we obtain the model-implied values of the IRS contracts in this work. We now review the adjustments we make to model (1) for the other contracts considered.

In the case of FRA, which are single-period contracts, we value a short (i.e. receive-fixed) position as the difference between the agreed-upon fixed-rate and the forward rate computed at valuation date. The rate difference is first multiplied by the notional and the length of time (in years) between the FRA effective date, which corresponds to the beginning of the forward-starting loan, and maturity date, which corresponds to the end of it, and then discounted back to valuation date.

In the case of OIS, the first adjustment that we make accounts for the fact that the floating leg pays the reference rate compounded daily over a set time period. This means that, whenever valuation date t_0 is past the effective date of the contract, the first floating leg payment is given in part by the compounded €STR rate that has prevailed in the market from the previous payment date to t_0, and in part by the €STR spot rate applicable to an investment that begins at t_0 and terminates at the next payment date. The second adjustment accounts for the fact that OIS payments are annual (i.e. at most one payment per year). This implies that, for spot-starting contracts, we compute the annual payments using 100% of the fixed and floating rates if maturity date is more than one year away from the contract effective date, otherwise we only use a fraction that is given by the number of months between the two dates, divided by 12. For forward-starting OIS contracts we always use 100% of rates to compute each of the annual payments.

EMIR article 9 data analytics software

Gabriel Grigorescu[a,1]

[a] *Oversight Department, National Bank of Romania, Bucharest, Romania*

Abstract. The National Bank of Romania (NBR)'s competences on EMIR apply only for investment services and activities provided by credit institutions established in Romania with derivatives based on currencies, debt, benchmarks and money market instruments. An analysis of EMIR data is necessary in order to assess supervised entities' compliance with the EMIR provisions, identify abnormal market behaviors and obtain a clear picture of the derivatives market. In order to fulfill this objective, a decision was adopted by the NBR to develop a dedicated software solution that uses EMIR to provide insight into the financial market and participants' behaviors. The solution was implemented in 2019 and continuously developed since then into a full technological ecosystem based on robotic process automation, business intelligence software and a database used to process and analyze the data as reported according to article 9 of EMIR and other financial data reportings.
Key words: NBR, robotic process automation, business intelligence, data-base.

1 Introduction

The European Market Infrastructure Regulation (EMIR) was introduced to mitigate the risks associated to derivatives market, increase transparency and provide a framework for prudential regulation and supervision of over-the-counter (OTC) derivatives market. Due to specificities of the Romanian domestic context, where there are two national competent authorities (the NBR and the Financial Supervision Authority), NBR is only responsible for investment services and activities, offered by credit institutions with derivatives based on currencies, debt, benchmarks and money market instruments. In order to ensure a proper implementation

[1] This publication should not be reported as representing the views of the National Bank of Romania. The views expressed are those of the authors and do not necessarily reflect those of the National Bank of Romania.

and enforcement of EMIR and to acquire the quality data needed to monitor the markets, the NBR found itself in need to evaluate its existing data processes and consider new innovative supervision tools.

The UiPath robot process automation software (RPA) was chosen to be the software solution capable of collecting and processing all the EMIR data, and importing them into a dedicated database. The RPA solution is powerful (it can process xml files up to 1gb), easy to use and configurable by NBR employees. For EMIR, the RPA is configured to be linked directly to each trade repository's database where it downloads the data, process them locally and uploads them to an internal database. The database has an interactive business intelligence interface, which is the only part of the process that the end user is able to see and that requires human input. The business intelligence interface provides updated indicators and information regarding every market participant, market shares, derivatives type, customer type, valuation throughout the life of a derivative contract, exposure for each participant, alerts, and data crosschecks. This integrated technology ecosystem is able to process not just EMIR data but also other financial reports that can be integrated with EMIR data and to expand the scope of the analyses. Since there are different technologies all working for the same purpose, the entire software solution was named MIS, an abbreviation for monitoring investment services.

2 EMIR data reporting approach

When an national competent authority (NCA) investigation is based solely on one data source, created as a result of transparency requirements set out for market participants and assessed ex-post, there might be a real possibility that, in case of market abuse or abnormal behavior, the reported data might be tampered with. Even if the data are corrected afterwards, the disruptive event already has occurred and the NCA's capability to contain the disruption immediately, might be reduced. In this context, in order to mitigate the risk of incomplete, incorrect and/or delayed reportings, we argue that an adequate level of data quality – necessary for monitoring financial markets – can be obtained through a holistic approach regarding investment services, financial markets and financial instruments. Within the NBR, such an approach has been achieved by custom developing a technological ecosystem based on three pillars: the RPA, the database and the business intelligence software. For this holistic approach, the RPA is used to correlate EMIR data with all sorts of other financial data acquired through Regulation (EU) 600/2014 (MIFIR), Regulation (EU) 2365/2015 (SFTR), Regulation (EU) 236/2012 and data from on-site investigations. Moreover, for financial stability purposes, the NBR has the statutory right to ask for and obtain statistical data regarding investment services and activities performed by the credit institutions.

Within the NBR, the view towards EMIR reporting is that the NCA must be able to adopt, when necessary, a fast response in order to contain and manage disruptive events; for this reason EMIR data reporting is administered directly by the oversight division, which was also designated to enforce the EMIR provisions. In this context, with the proper tools, the data provided via EMIR can be used successfully for other purposes as well, for example to detect abnormal market behaviors that might indicate benchmark manipulation, short selling, or other market abuses.

3 MSI solution - RPA

The purpose of the RPA is to automate business processes, mainly repetitive tasks such as data input, processing and manipulation. Considering the EMIR data format, the best RPA for this project proved to be a software solution provided by the UiPath company. For the purpose of this paper it should be noted that the NBR collects data directly from the trade repositories (TR) as per EMIR article 81. Prior to the EMIR Refit, the standard EMIR data format was subject to a different interpretation by each TR and even if the fields had already been established by the regulatory technical standards, it was noted that while one TR provided the data as an xml file another opted for a csv file; there were also minor differences in the standard fields from TR to TR; for example a simple space in the field name could prove to be a challenge in the data correlation and import. In this context, a fundamental first function of the RPA was to harmonize the data of all TRs, by identifying each field and importing it into the database in the same format. After the EMIR Refit, when there was a full standardization of TR data, the RPA harmonization function was redesigned to identify, by LEI codes, each entity under NBR supervision regardless of their position (reporting or other counterparty).

The second fundamental function of the RPA is to correlate the EMIR data with MIFIR and a national statistical data reporting platform that also includes granular data about OTC and traded on trading venues (TOTV) derivatives (SIR-BNR). As such, the robot can make connections between every field and it's equivalent (when available). Moreover, the RPA was taught to recognize all the details of a reported transaction and to perform a crosscheck among data sources to identify if that EMIR transaction was also reported via SIR-BNR or if the same transaction was reported in accordance with EMIR by both counterparties. It is expected that every transaction should be found two times in the TRs aggregated data, one time where the supervised credit institution acted as reporting counterparty and another in the position of the other counterparty. For the crosscheck feature, the RPA works only in connection with the business intelligence interface. Moreover, the RPA, using external data sources, is also able to determine if the same transaction had different details reported by each counterparty.

The third fundamental function of the RPA is to check every financial instrument and include it in a library, thus enabling a better understanding of OTC and TOTV derivatives (see Figure 1).

4 MSI database

For the purposes of storing the EMIR data, NBR uses a custom made database based on the Microsoft SQL relational database management system. This software was chosen for its versatility and support of a variety of business intelligence and data analytics applications. The structured query language permits administrators of MSI to perform additional verifications to ensure every process of the RPA

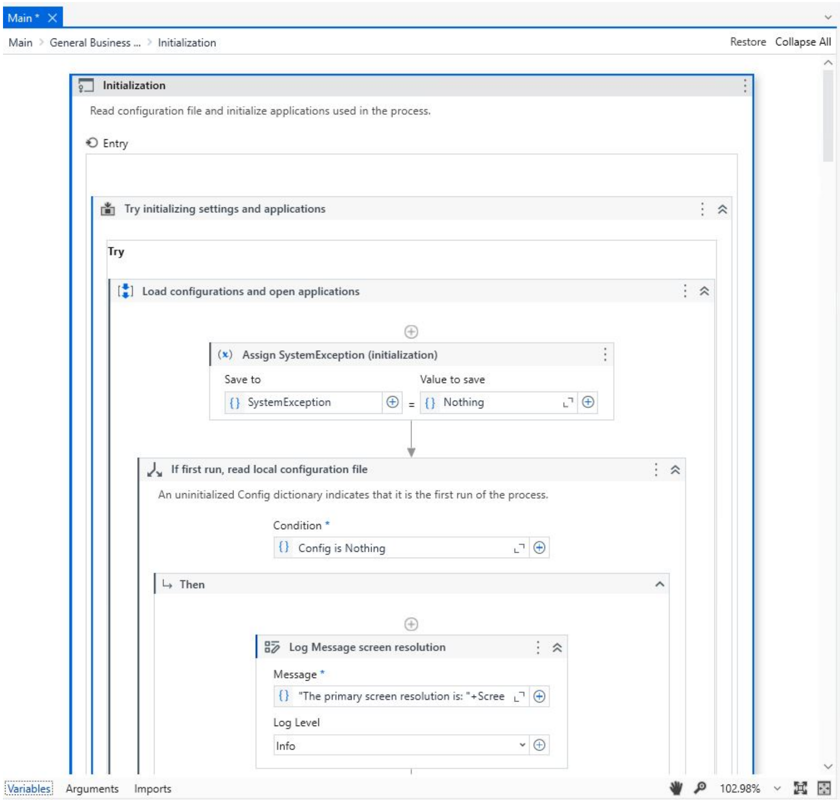


Figure 1 – MSI UiPath main, altered for demonstration purposes.

and business intelligence is working properly. The architecture of the database includes tables designed to store all the data sources, for example, EMIR, MIFIR, SFTR, and SIR, as well as an entire library of data regarding financial instruments, counterparties, currencies and exchange rates, MIC codes, investment services and activities authorizations. The library is used mainly to teach the RPA different functions and prepare for the future implementation of artificial intelligence technology. In order to avoid refresh problems or extended loading time, the database has strict rules to refresh the data overnight, source by source, at different times outside business hours. Furthermore, the entire architecture of the database was designed to be compatible and integrated with a business intelligence solution.

5 MSI business intelligence

After solving the challenges of gathering, processing and storing the data, the NBR needed to find a suitable tool capable to deliver insights for EMIR data via interactive dashboards. To fulfill the oversight requirements, the business intelligence solution was configured with 4 different dashboards.

5.1 Main dashboard

The MSI main dashboard provides an overview of the entire market based on the total number and value of concluded transactions (see Figure 2). The user is able to filter the data by selecting the timestamp, credit institutions, types of transactions and financial instruments, market share and counterparty type. Moreover, since this solution covers multiple data sources, not just EMIR, the user has the possibility to choose the data source. This dashboard is useful when assessing concentration risk, large exposures per participant, preferred types of financial instruments, and types of counterparties. Since EMIR transactions are reported with different currencies, before displaying the visuals MSI translates the reported currencies using the official exchange rates in order to calculate and display the values in romanian lei (RON), euro (EUR) or US dollars (USD). The NBR calculates daily exchanges for RON with 36 other currencies. If the user wants to see the situation in RON and multiple transactions are reported in currencies other than the 36 available, the conversion is made to EUR or USD and afterwards to RON.

Figure 2 – Main dashboard.

5.2 Comparative dashboard

The MSI comparative dashboard performs an in-depth comparative analysis between two supervised credit institutions for multiple periods of time (daily, weekly, monthly, yearly) or a comparative analysis of one credit institution's performance against the market (see Figure 3). For EMIR, the comparative dashboard shows the value considering all the valuations, updates and modifications of a contract at the selected date. The user is also able to asses the performance of multiple participants in specific derivatives. For example if a bank presents a sharp increase in its total number of FX derivatives compared to the bank's activity in a previous timestamp or compared to the activity of other major market participants, a red flag appears; further inquires can then be carried out to determine if the anomaly was the result of abnormal behaviors or if the situation had other justified explanations. The indicators are overlapped for each period to enable a parallel view based on the total number of transactions and the total value. Another feature of the comparative dashboard is the financial instruments overview; here, the user is able to configure the view to display data for one, many or all financial instruments throughout the selected time period.

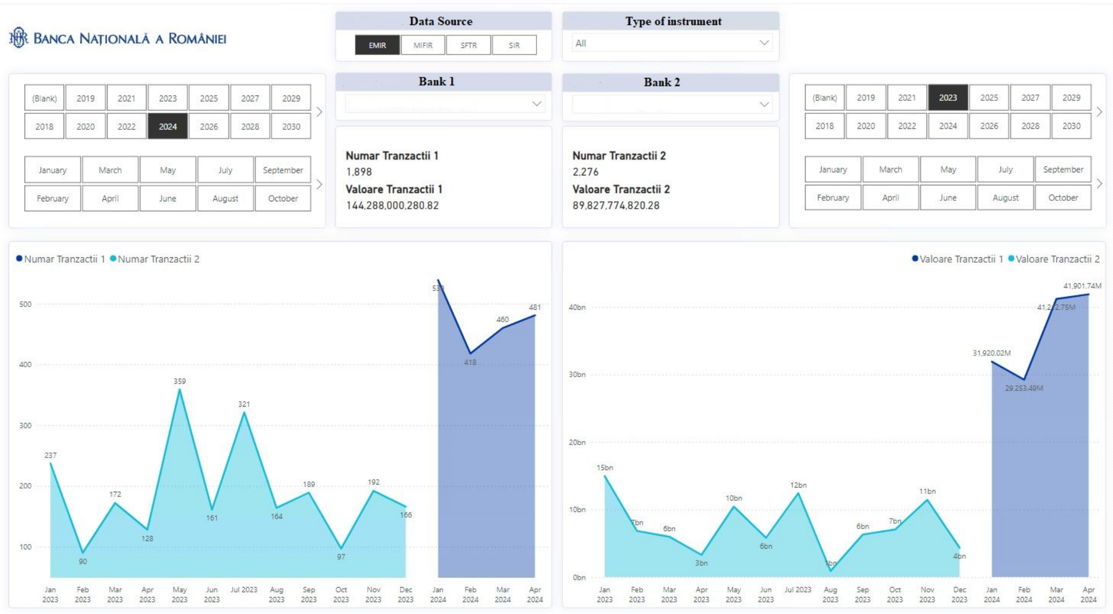


Figure 3 – Comparative dashboard.

5.3 Crosscheck dashboard

One of the biggest challenges in employing oversight software as an NCA is to ensure an adequate level of data quality. The NBR only can access EMIR data, without the possibility to check in real-time if a transaction's details are correct at the reporting date. In practice, it was noticed that many transactions had been corrected after they were reported, even for important fields such as contract values and quantity. Before MSI, the only feasible way to verify the correctness and completeness of reported transactions was to perform on-site investigations on specific samples of data. One solution to raise the data quality level was to perform an automatic crosscheck between EMIR data and SIR-BNR whih collects a large range of financial data for financial stability assessments, including on derivatives. In order to perform this crosscheck, MSI analyzes if every EMIR transaction matches an SIR-BNR transactions. In order for a transaction to match, contract's date and time, counterparties, value, exchange rates and instrument fields must match. In practice, it was observed that not all the transactions match especially due to different reporting dates. Analyzing this anomaly, it was found that the reporting entities were able to provide valid arguments as to why the reporting date differed for the EMIR and SIR-BNR transactions. One commonly-shared explanation was that the account used to upload SIR-BNR transactions had expired and the NBR had not renewed it in time; some explanations involved other SIR-BNR connectivity issues. Therefore, it was decided that an EMIR/SIR-BNR crosscheck was not enough to ensure a comfortable level of data quality, and a new crosscheck was employed. Considering that the NBR receives the data directly from the TRs, certain queries were developed so as to request all data where the

LEI code of entities under the NBR's supervision appears. By employing this filter, it was observed that every transaction must be found two times as reported by each counterparty especially when different TRs are used. As a result, MSI is able to find every transaction as reported by each counterparty. After performing these two crosschecks, the oversight team is able to identify if a transaction was reported in time and if all the details match. In practice, more than 90 percent of the transactions are matched. For the remaining transactions, the oversight department performs on-site investigations, issues recommendations and deadlines and, when necessary, adopts punitive measures.

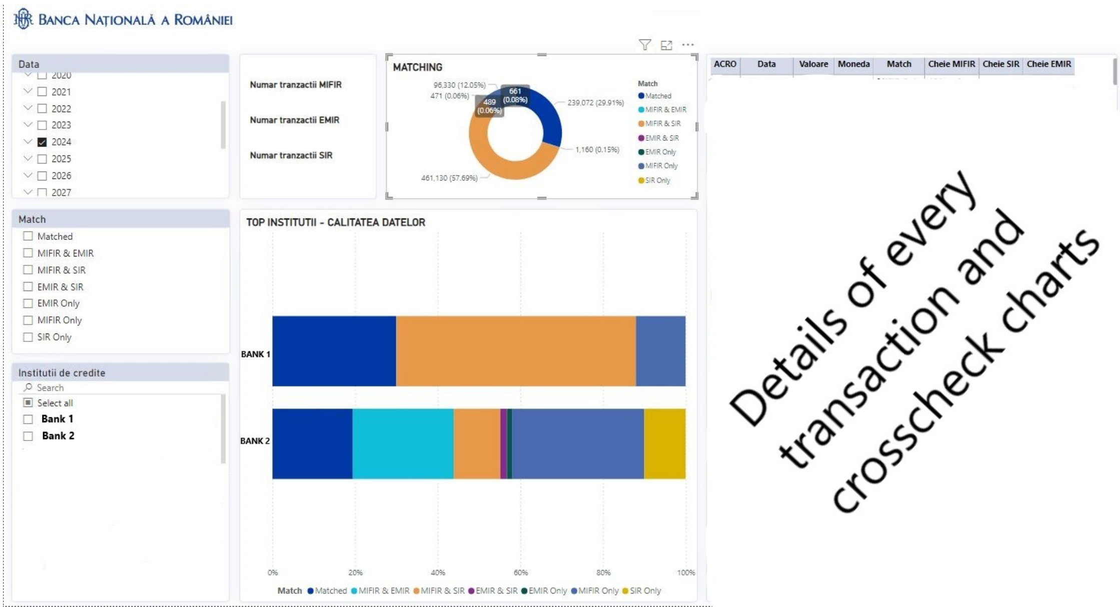


Figure 4 – Crosscheck dashboard *for confidentiality reasons data and charts are removed.

5.4 Alerts dashboard

When designing the MSI architecture and functionalities, it was noticed that the MSI solution could successfully provide meaningful insight for other oversight tasks too. In addition, it can be a trustworthy indicator for abnormal market behaviors and even a market abuse indicator when certain thresholds are exceeded. Therefore, an MSI algorithm was developed that identifies when a reporting entity is close to or exceeds the mean value of transactions for specific instruments and classes of instruments for the past three months and automatically displays a visual alert and sends an email to oversight managers. The algorithm also permits user input for the threshold calculations, thus enabling the user to set alerts for specific instruments and periods of time, as well as for specific counterparties in order to to provide alerts when certain transaction levels between specific counterparties are exceeded (see Figure 5).

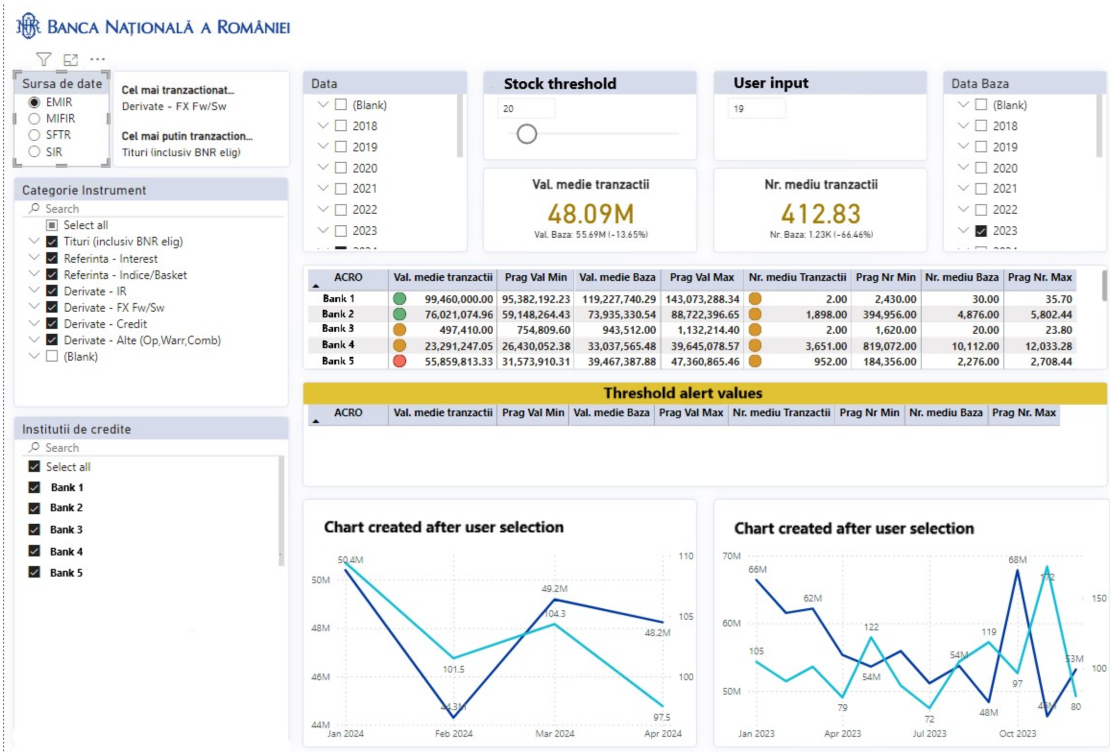


Figure 5 – Alerts dashboard *for confidentiality reasons data and charts are removed

6 Conclusions

Every central bank and NCA owns and manages huge amounts of data and sometimes new technologies can facilitate deeper, more meaningful insight into these data. If ex-ante detection is a utopia, the NBR's target is to reach a qualitative assessment as close to real-time as possible. Therefore, the future of MSI solutions is in artificial intelligence and machine learning, where new technologies can enhance MSI capabilities to detect patterns and behaviors.

Implementation of EMIR Refit reporting at Banque de France: technical challenges and insurance supervision use case

François Pégory[a,1] and Michele D'Ambrosio[a]

[a] *Research and Risk Analysis Directorate, ACPR, Paris, France*

Abstract The objective of this work is twofold. First, it shows how Banque de France takes on the technical challenges of receiving regulatory data on derivatives transactions under EMIR Refit regulation. Then, it presents a use case of such data from the standpoint of the macro prudential supervision of insurance undertakings in France, combining the use of EMIR data with regulatory reporting under Solvency II directive.

Key words: Hadoop, EMIR Refit, financial derivatives, Solvency II, insurance supervision.

1 Introduction

The European Market Infrastructure Regulation (EMIR)[2] published in 2012 introduced the requirement to report all transactions in derivative contracts in Europe. The reported data are very rich, including, among others, the contract and counterparty identifiers, the notional value and market value, and other data points on technical features of the contract and the underlying. Reportings are centrally collected by trade repositories (TR), which are private companies supervised by the European Securities and Markets Authority (ESMA); the TRs act as authoritative registries of key information regarding open over-the-country (OTC) derivative

[1] The views expressed in this publication are those of the authors and do not necessarily reflect those of the ACPR.

[2] 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

trades, providing an effective tool to mitigate the inherent opacity of this market. In 2019, EMIR was amended through a Regulatory Fitness and Performance Program (EMIR Refit). After consultation, ESMA published new regulatory[3] and implementing[4] technical standards. The new taxonomy is meant to improve the quality of the reporting and is in line with the guidance of the International Organization of Securities Commissions (IOSCO) with respect to the introduction of a new unique product identifier (UTI). Ensuring clear rules on how product identifiers are created is key to ensure reconciliation of records.

Following the introduction of Refit, in a move that strengthens regulatory convergence, the European Insurance and Occupational Pension Authority (EIOPA) included the possibility to declare the same unique trade identifier of EMIR contracts under Solvency 2 Quantitative Reporting Templates (QRT) on derivatives - called S.08.01 - starting with taxonomy 2.8.0 effective from December 2023.

2 From data source to raw tables

In what follows, the main challenges faced with the introduction of Refit are presented.

Under EMIR Refit, Banque de France receives more than 15 million data rows (concerning France only) per day from the four trade repositories licensed in the EU via an exchange platform mangaged by ESMA. In order to accommodate for this large amount of data, a new integrated data platform relying on an Hadoop ecosystem was set up at the Banque. This allows users to manage, analyse and extract value from large volumes of data using PySpark as the main programming language.

The new Refit reporting format is the standardised ISO-20022-XML, which is hierarchical: therefore, in order to build a table out of the data files ingested, we need to leverage on Hive table technology. We will call these output tables "raw tables", as they contain the information "as-is", i.e., with no modifications or corrections with respect to submissions from reporting entities. With Apache Airflow, a workflow management platform for data engineering pipelines,[5] an automatic orchestration of downloads is scheduled on a regular basis.

In addition, in order to stay close to the data source, we transform the xpath of an attribute to the column name by replacing the character '/' with '_'. For example, for the LEI of the reporting entity, the xml xpath will look like the following:

- 'CtrPtySpcfcData/CtrPty/RptgCtrPty/Id/Lgl/Id/LEI' ;

---

3 Commission Delegated Regulation (EU) 2022/1855 of 10 June 2022 supplementing Regulation (EU) No 648/2012 of the European Parliament and of the Council with regard to regulatory technical standards specifying the minimum details of the data to be reported to trade repositories and the type of reports to be used.

4 Commission Implementing Regulation (EU) 2022/1860 of 10 June 2022 laying down implementing technical standards for the application of Regulation (EU) No 648/2012 of the European Parliament and of the Council with regard to the standards, formats, frequency and methods and arrangements for reporting.

5 See https://airflow.apache.org.

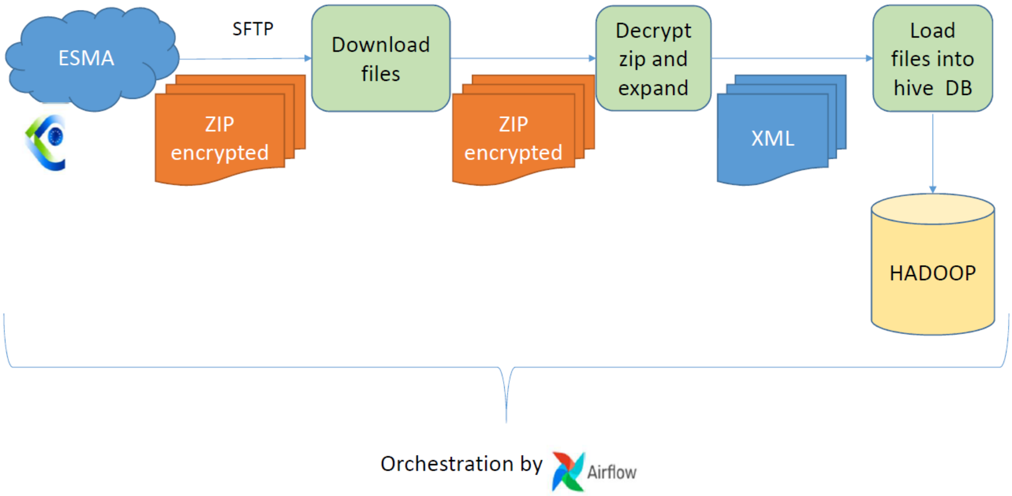


Figure 1 – Data pipeline: from raw to enhanced tables.

Therefore, the resulting column name will be:

- 'CtrPtySpcfcData_CtrPty_RptgCtrPty_Id_Lgl_Id_LEI'.

Data are daily reported across four main types of file: all outstanding contracts at the end of the day (*state* data), transactions occurred during the day (*activity* data) and margin calls data (both for *state* and *activity* data). For each one of these files, one table is created by mapping each attribute to a single column. In case of multiple tags for a single attribute, an array structure is used to collect all information together. For instance, if the counterparty type of a company is *"bank"* and *"insurance"*, the corresponding unique column in the data table will be { *"bank"*, *"insurance"*}. Some data are also weekly reported about rejections, warnings and reconciliations of transactions. For theses files, some attributes can even be more complex, as they can repeat and have children. This calls for more sophisticated mapping structures. In this case, we split information across two databases - *main* and *detail*: in the latter, the most complex information is collected and is made available for users to exploit.

Lastly, some useful data points are not declared under EMIR and need to be fetched from other sources (see Figure 1). For instance, the attribute 'country of the reporting counterparty' is retrieved from GLEIF based on entities LEI; further information on the international securities identification number (ISIN) and unique product identifier (UPI) of the contracts are pulled out of ANNADSB database, which is managed by the Derivative Standard Bureau (DSB).[6]

[6] See https://www.anna-dsb.com/.

3 Enriched tables: an answer to raw data tables challenges

In order to overcome the different challenges presented above, a number of different solutions are put in place.

In order for business users to exploit the very large data tables described above, a solution based on PySpark was implemented. This language was already mastered by IT developers, enabling them to train business users for enhanced efficiency.

Hive query system allows to query data within an array, therefore mastering the complexity of the raw tables. Consider the following query:

```
Select count(*) as count
from emir_state
where array_contains(reporting_nature, ''insurance'') and
    ref_date =''2024-09-10'';
```

This amounts to count the number of rows where the 'reporting counterparty' is at least an insurance. Despite the complexity of the data structure of the column 'reporting nature', applying a filter becomes incredibly simple thanks to the *array_contains* feature.

The complexity of the raw table is also reflected by its size: they can have somewhere from 300 up to 700 columns. To make data query easier, we decided to create multiple tables instead of one single *state* table based on asset classes: in this way, only relevant columns for each of the six asset classes are selected.

With the introduction of EMIR Refit, some columns have been renamed, and some others have changed their nature. Consider the following case: the LEI of the reporting counterparty used to be under the following tag

- 'ctrptyspcfcdata_ctrpty_rptgctrpty_id_lei'

the same information is now under another tag

- 'ctrptyspcfcdata_ctrpty_rptgctrpty_id_Lgl_id_lei'.

To ensure compatibility across the two taxonomies, we established that common names be used in the *enriched* tables: in this example, the name chosen was 'reporting_lei'. This also reduces the entry cost of business users when approaching data tables as it makes clear what to expect from each column.

Lastly, to enhance the quality of the information available to users, we implement data collection of the characteristics of OTC contracts from the ANNADSB database. For each one of our enriched tables by asset class, information from ANNADSB are added via a merge operation. The GLEIF database is also used to retrieve the 'country of the reporting counterparty'.

The flowchart in Figure 2 shows the pipeline that leads to the creation of the enriched tables starting from raw data and adding external data sources. As a result, six tables are available to users in a format that is easier to access than the original one.

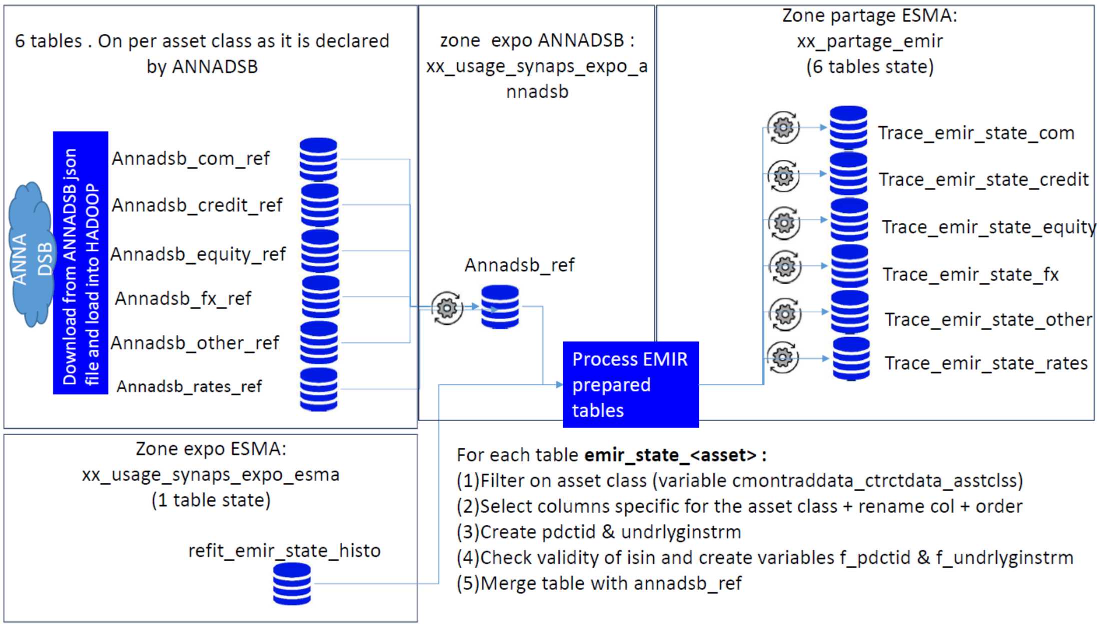


Figure 2 – Data pipeline: from raw to enhanced tables.

4 Data selection and filtering

We provide an overview of the data used in the next chapters. Solvency 2 data come from QRT on two reporting dates: December 31, 2023 and March 31,2024. Data are then extracted from EMIR on a biweekly basis, starting from September 30, 2023 to March 31, 2024. The reporting population includes all insurers who reported the QRT S.08.01 over the observed period. We take advantage of the dual-sided reporting under EMIR, i.e. both counterparties to derivative contracts are required to report if they fall under the scope of the EMIR regulation. Thanks to this feature, the reported leg of an insurer's transaction can be compared with its mirroring leg to ensure that key variables such as contract type, asset class, notional value and currency are matching across both reporting. If so, then we keep the insurer's record; elsewise, we drop it.

We also develop a simple data quality indicator. If any of the following is true, the indicator will flag an issue:

- maturity and execution dates are Null;
- notional currency is Null;
- market value currency is Null;
- contract type is Null;
- asset class is Null;
- UTI is duplicate.

5 Convergence in data reporting between EMIR and Solvency II

The update of QRT S.08.01, which entered into force in late 2023[7] introduced reporting of the unique trade identifier (UTI) as the contract identifier. This allows for simpler comparison of reported data at a very granular level.[8] In addition, the latest technical specifications[9] impose that the field 'notional value of the derivative' be reported in the original currency. As the previous version of the text[10] was silent on the reporting currency, one would assume that reporting was done in euro. Nevertheless, we observe several inconsistencies in the reported figures at entity level under the two texts, potentially implying diverging interpretations across reporting entities.

Figure 3 shows the extent to which reconciliation was possible over the first two quarters where UTIs were available in Solvency 2. While it appears that a different narrative applies to each asset class, some recurring patterns were observed:

- as is the case for credit and foreign exchange derivatives, between 10% and 50% of the UTIs were declared in both databases, with comparable figures between Solvency 2 and EMIR;
- for equity and interest rate derivatives we observe instead that the rate of shared identifiers diverges greatly with each asset class; if shared UTIs account for 15% to 30% in EMIR across the two asset classes, this range is much higher (and larger) in Solvency 2, reaching 35% to 70%.

The variability observed in the rate of shared UTIs across asset classes is somewhat disappointing. We hoped for a fairly high rate of shared identifiers consistently across all asset classes, which instead is not observed. This constitutes nevertheless a solid point for developing further analysis over what may cause such variability across asset classes and time periods. We will further investigate into the non-shared records to assess to what extent the low pairing ratio is due to data quality issues.

6 EMIR data for supervisory purposes

The high frequency and granularity of EMIR data are very appealing features for supervisors, all the more so when these features are compared with those of Solvency 2.

While any updates to a position in EMIR is reported on a daily basis, so that at T+2 updated data are available, Solvency 2 data are reported on a quarterly

[7] EIOPA XBRL Taxonomy 2.8.0, see https://www.eiopa.europa.eu/tools-and-data/supervisory-reporting-dpm-and-xbrl_en.

[8] It should be noted that certain derivatives are exempted from reporting under EMIR, such as, for example, some intragroup transactions.

[9] Commission implementing regulation 2023/849 on reporting and disclosure, http://data.europa.eu/eli/reg_impl/2023/894/oj.

[10] Commission implementing regulation 2015/2450, http://data.europa.eu/eli/reg_impl/2015/2450/oj.

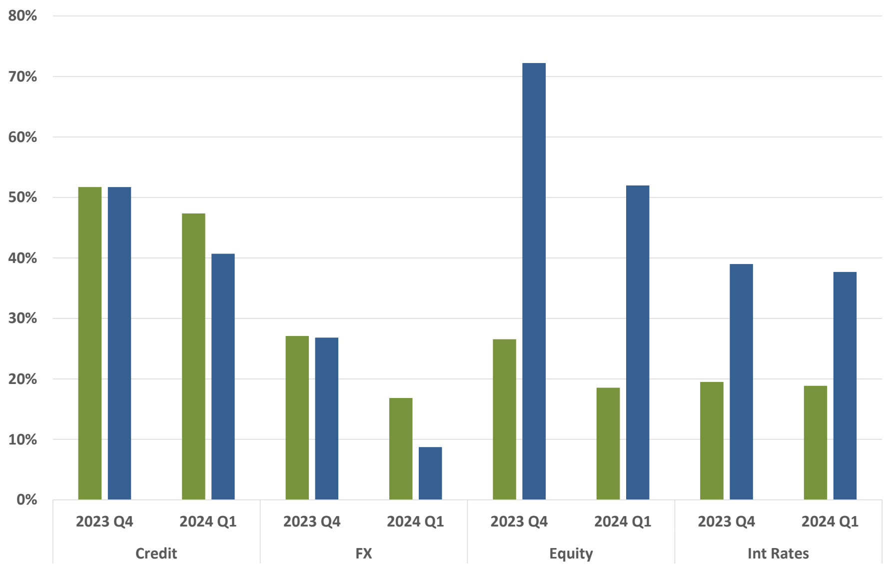


Figure 3 – EMIR and Solvency 2 shared UTI per asset class.

basis, with quarterly reporting being available about two months later than the date to which they refer. In addition, EMIR data are richer than Solvency 2 as they include data on margin calls and specific information on the contract and the underlying, leaving room to the possibility of analysing the interconnectedness of the system, including companies other than insurers.

Figure 4 shows the notional value declared in the selected EMIR data over a six-month period at a bi-weekly frequency. Thanks to the high frequency of the data, one can observe the aggregate notional amount peak at year-end 2023 and the subsequent drop, slightly above mid-November levels.

On the other hand, the two lines (in blue: notional value of contracts with shared UTI with Solvency 2 over total; in red: all contracts with data quality issues) stay mostly flat over all the thirteen periods observed. This likely means that the contract population reported under Solvency 2 is less volatile than EMIR. In addition, we speculate that the increase in total outstanding notional along with weak movements of the shared notional curve is motivated by short term contracts that are not declared under Solvency 2. This hypothesis will be tested in the future.

7 Conclusions and next steps

Pre-Refit data and Refit data are now available to Banque de France users in separate and harmonised tables. Analysing how each attribute changes with the new taxonomy (e.g., name of the attributes, possible values, format of the reporting) is key to make access to the data easier for business users.

Another way of improving the data quality is to get UPI from ANNADSB to enrich Refit data. The UPI identifier was introduced under Refit. Only some contracts[11] have ISINs; otherwise, contracts are supposed to have a UPI that allows

[11] If the product is admitted to trading or traded on a trading venue or traded on a systematic

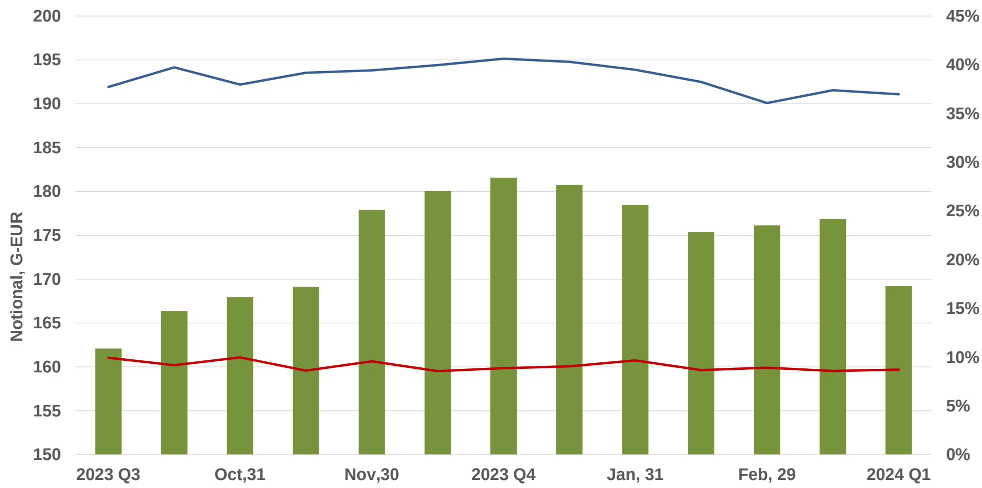


Figure 4 – EMIR notional value of contracts with **shared** UTI in absolute terms (LHS) and relative terms (RHS). Data quality indicator: All contracts (shared, non shared) with data quality issues (LHS).

for describing their main features. Therefore, we expect the use of UPI to increase going forward.

On the data analysis side, the first step is to understand what motivates the lower-than-expected ratio of shared contract identifiers between the two databases. However, an analysis of more granular data that is not shown in this work hints to encouraging results: the breakdown of the aggregate notional broken down by contract type and date is very much similar across the two data sources. While this point has not been developed in this paper, the comparison of aggregate metrics (beyond those concerning UTIs) across data sources will allow for finer cross-validation of records for each reporting entity.

Leveraging data on margin calls in EMIR can also give new insights on credit and liquidity risk. Overall, using EMIR data can give an edge on the timeliness and scope of the analysis of several market risks. A visual representation of interconnections (e.g., networks) of market participants is another step towards a more extensive use of EMIR data.

Lastly, offering users access to data through a business intelligence tool, either for data quality purposes or for supervisory monitoring, could simplify access and enhance both data availability and oversight.

internaliser and the underlying is admitted to trading or traded on a trading venue, or is an index or basket composed of instruments traded on a trading venue.

Mining equity implied volatilities from EMIR

Michele Leonardo Bianchi[a,1] and Daniele Miceli[b,2]

[a]*Financial Stability Directorate, Banca d'Italia, Rome, Italy*
[b]*Directorate General Macroprudential Policy and Financial Stability, European Central Bank, Frankfurt am Main, Germany*

Abstract. In this work we propose a framework to extract implied volatilities from option prices reported in the EMIR database. We compare these estimated volatilities with the surfaces obtained by third-party data providers and we conduct advanced data quality checks on these volatilities. Starting from option prices, we show how it is possible to estimate implied volatilities surfaces and to assess the goodness of these estimates. We describe the EMIR equity option data reported between January 2021 and April 2024 along with empirical results for four major underlyings on June 30, 2023.

Key words: EMIR data, equity options, implied dividend yield, implied volatility.

1 The authors are grateful to Arianna Miglietta, Salvatore Cardillo, Dario Ruzzi and Gwyneth Schaefer for their helpful suggestions. This publication should not be reported as representing the views of the Banca d'Italia. The views expressed are those of the authors and do not necessarily reflect those of the Banca d'Italia.

2 Daniele Miceli attended an extracurricular internship program promoted by Sapienza University at the Banca d'Italia during the period from February to July 2024. The work was conducted as part of this internship.

1 Introduction

With the aim of improving stability, transparency and efficiency in derivatives markets, on 16 August 2012, Regulation (EU) No 648/2012, better known as the European Market Infrastructure Regulation (EMIR), entered into force. Under EMIR, market counterparties holding any form of derivative are obligated to report data on derivative contracts to a trade repository (TR) recognized by the European Securities and Markets Authority (ESMA). Transaction-by-transaction EMIR data on derivatives are an extremely rich source to extract information on the behavior of financial market participants, which is useful to build financial stability monitoring tools (Agostoni et al. [2024] and Bianchi et al. [2025]). The significant advantage of working with EMIR data lies in its basis on real transactions where counterparties are identified. These data can be explored from different perspectives, since there are derivatives of various types and with heterogeneous underlyings. In this work we focus on options written on equities and equity indexes with the aim to extract from option prices the so called implied volatilities (IVs). Though this exercise is straightforward in theory, it becomes challenging when dealing with millions of prices.

In option markets the implied volatility (IV) is the key parameter to determine the price of a derivative contract and option prices are often quoted in terms of this volatility; this does not occur in EMIR. This metric represents the standard deviation to be inserted into the Black-Scholes pricing formula to match the price of the option. Under the Black-Scholes pricing model, IVs vary depending on the strike price and the maturity of the option (see Section 4). Typically, they form a pattern known as the volatility smile or smirk, which resembles, for a given maturity, a U shape, not necessarily symmetric. The wings of this smile extend towards the deep in-the-money (ITM) and out-of-the-money (OTM) strikes. Specifically, the left wing tends to be steeper than the right, and there is usually a minimum point around the at-the-money (ATM) strike price.

IVs, as extracted from option prices, reflect investors expectations about future price movements. They provide insights into market sentiment; they help investors optimize and hedge the risks of portfolios (i.e. there are derivatives where the payoff explicitly depends on some measure of volatility); and they are the main input for the pricing and risk management of other derivative instruments. For these reasons, IVs play a significant role in economic and financial analysis by offering a real-time indicator of market sentiment, risk, and uncertainty. They provide insights into market participants' perspectives on broader economic conditions and help identify periods of mispricing or excessive optimism or pessimism in the markets. Unlike historical volatilities, which look backward at past price fluctuations, IVs provide a forward-looking measure of how much market participants expect an asset price to move in the future; they also can be used to capture market confidence or fear surrounding economic events. For example the CBOE Volatility Index (VIX), also known as the Fear Index, is derived from the prices of the options having as underlying the S&P 500 Index. A rise in IVs across equity or bond markets often signals increased uncertainty or concerns about recessions, inflation, or interest rate changes. A higher level of IVs often indicates apprehension about impending financial or geopolitical shocks. Additionally, changes in IVs can offer insights into

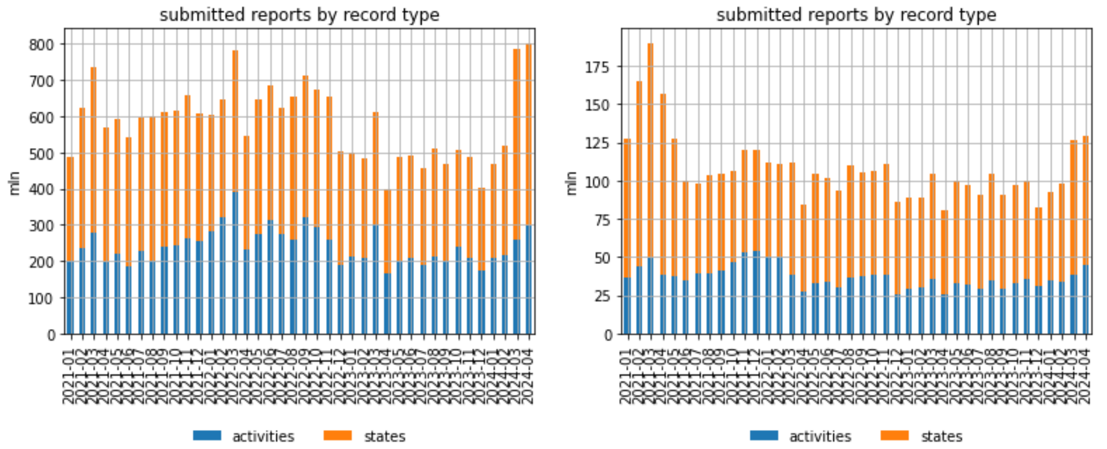


Figure 1 – EMIR data available at the Banca d'Italia: all derivatives contracts (lhs) and equity options (rhs). Number of monthly reports by record type. The underreporting detected in 2023 was solved as of March 2024. This results in an increase in the volume of collected data in the first quarter of 2024.

business cycles, as volatility tends to be lower during periods of economic expansion and higher during downturns, making it a potential leading indicator of economic shifts.

This work provides a detailed exploration of EMIR data collected by Banca d'Italia on equity options and the technical details for the extraction of IVs. After having discussed the implementation of additional data-tailored filters, we outline a method to extract IVs from option prices, emphasizing the critical role of model inputs, such as the dividend yield. Advanced data quality checks are highlighted throughout to ensure the reliability of estimated IVs. We show that EMIR data enables authorities to independently estimate their own implied volatilities on real transaction data and across a wide range of underlying assets and counterparties, removing the necessity of relying on external financial data providers for these IVs. We conclude by suggesting potential further analyses.

2 EMIR data on equity options

Banca d'Italia collected in the period from January 2021 to April 2024[3] more than 23.4 billion EMIR data points, i.e. over 580 million data points each month (see Figure 1). Around a fifth of this huge amount of data is related to equity options.[4] In the analyzed period, Banca d'Italia received almost 4.3 billion data points related to equity options. After cleaning and deduplicating the data, we have a total of 2.2 billion and 1.7 billion data points, respectively (see Bianchi et al., 2024, for a description of the preprocessing steps).

In EMIR there are trade activity reports to which we refer to as activities ('A'),

[3] The reporting framework currently into force (EMIR refit) went live on 29 April 2024. This analysis is based on the data collected under the previous reporting framework.

[4] According to EMIR, equity options are reported as derivatives transactions with the asset class field equal to 'EQ' and contract type equal to 'OP'.

record type	option type	option style	n.obs	n.obs (%)
S	C	A	650,458,778	37.828%
S	P	A	643,117,211	37.401%
S	P	E	182,366,435	10.606%
S	C	E	124,481,256	7.239%
A	P	E	34,940,222	2.032%
A	C	E	30,877,140	1.796%
A	C	A	22,713,153	1.321%
A	P	A	22,397,845	1.303%
S	C	B	7,860,631	0.457%
S	O	E	127,621	0.007%
S	P	B	99,135	0.006%
A	O	E	19,684	0.001%
S	C	E,A	14,670	0.001%
A	C	B	11,574	0.001%
A	P	B	11,048	0.001%
S	O	B	2,095	0.000%
S	NaN	NaN	969	0.000%
S	O	A	617	0.000%
S	C	NaN	588	0.000%
S	P	E,A	131	0.000%
A	O	A	30	0.000%
A	O	B	12	0.000%
total			1,719,500,848	100.0%

Table 1 – EMIR data available at the Banca d'Italia, after cleaning and deduplication. Record type: A (activities) and S (states). Option type: C (call), P (put) and O (other). Option style: A (American), B (Bermudian), E (European). When the information is not available, NaN is reported. Equity option data between January 2021 and April 2024.

containing all the reports sent the previous day, and there are trade state reports, which we refer to as states ('S'), containing all pending trades at the end of the reporting day. While activities are related to the lifecycle events of a transaction (e.g. new transaction, conclusion, valuation, modification, termination), states represent the outstanding transactions at the end of the day. These outstanding transactions are usually considered for risk monitoring purposes.

In Table 1 we report some summary statistics across option style (i.e. call, put or other), record (i.e. activities or states) and option type (i.e. European, American or Bermudian). More than three quarters (76%) of the total observations refer to the states of American options ('A') and only a small fraction (18%) are the states of European options ('E'). Bermudian options, representing less than 1% of the total observations, will not be considered in the empirical analysis.

Since the underlying identification is not necessarily an ISIN code, a mapping is needed to connect the information provided in the EMIR field with the LSEG data on both underlying and option data. A manual matching step allows us to establish connections between a given underlying asset and its various representations in EMIR. Through this process, we are able to identify approximately 1,800 distinct underlying assets from the pool of over 110,000 reported identification strings. Furthermore, we are able to obtain from LSEG both price and IV data for approximately 1,700 of these underlying assets (we have a total of almost 1.5 billion EMIR data points). As depicted in Table 2, there is a significant dispersion among various underlying assets, with the top 15 accounting for only 24% of the total data points.

LSEG ticker	name	option symbol	type	n.obs	n.obs (%)
DJES50I	Euro Stoxx 50	GXE	index	61,014,801	4.1%
DAXINDX	DAX	DAX	index	37,098,427	2.5%
S&PCOMP	S&P 500 Composite	SPX	index	33,049,180	2.2%
AMSTEOE	AEX	EOE	index	32,483,998	2.2%
GB00BP6MXD84	Shell	RDAS	equity	24,894,710	1.7%
LU1598757687	Arcelormittal	AMMT	equity	21,635,540	1.5%
NL0011794037	Koninklijke Ahold Delhaize	HAH9	equity	20,924,590	1.4%
JAPDOWA	Nikkei 225 Stock Average	NK2	index	20,323,486	1.4%
HNGKNGI	Hang Seng	HSII	index	19,960,359	1.3%
NL0011821202	ING Groep	ING	equity	16,982,819	1.1%
NL0010273215	ASML Holding	ASML	equity	14,088,540	1.0%
NL0000009538	Philips Electronics Koninklijke	PHIS	equity	12,287,624	0.8%
US78462F1030	SPDR S&P 500 ETF Trust	SPY	ETF	11,789,714	0.8%
FTSEMIB	FTSEMIB	SMIB	index	11,609,532	0.8%
DE000BAY0017	Bayer	BAYR	equity	11,021,403	0.7%
other	-	-	-	1,130,071,585	76.4%
total				1,479,236,308	100.0%

Table 2 – EMIR data available at the Banca d'Italia, after cleaning, deduplication and mapping with LSEG data. We report the LSEG ticker of the underlying, its name and type, and the option symbol needed to obtain option data from the provider. Equity option data between January 2021 and April 2024.

3 Extracting implied volatilities from option prices

This section describes the framework developed to compute IVs from EMIR equity option prices. First, we define the factors needed to estimate IVs, including the risk-free interest rate and the dividend yields related to each single underlying. Second, we perform an additional data quality step tailored on equity option data. Then, we describe the methodology to compute IVs and show the empirical results for two indexes, the Euro Stoxx 50 (DJES50I) and the Ftse Mib (FTSEMIB), which we refer to as EuroStoxx and FtseMib, and two stocks, Shell (with ISIN code GB00BP6MXD84) and Unicredit (IT0005239360). We then compare our estimates with those obtained from LSEG. Lastly, we propose our methodology to flag outliers and to smooth EMIR equity IV surfaces. The algorithms presented in this work have been implemented using Python.

3.1 Model inputs

We select a valuation date (June 30, 2023) and an underlying (a stock or an index) having options traded and reported in EMIR at the selected valuation date. EMIR provides equity option quantitative data, such as the strike price and the price of the option,[5] and contract characteristics such as maturity, type, style, information on the underlying (the name or the corresponding ISIN code) and information on the product (e.g., the ISIN code of the option, if it is exchange traded). As risk-free rate we consider the euro short-term rate (ESTR) overnight index swap (OIS) rate provided by LSEG, together with the underlying price value. LSEG provides for ESTR OIS maturities up to thirty years, as shown in Figure 2. It is important to note that most of the options have maturity less than ten years. The underlying spot prices are also obtained from LSEG.

[5] There are both reference currency and euro denominated data.

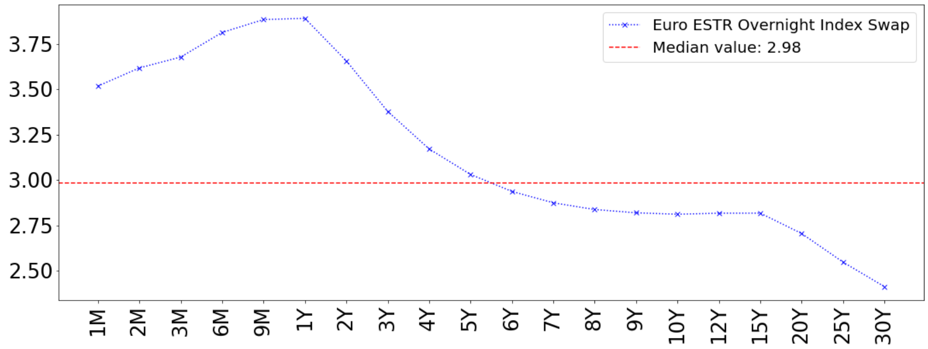


Figure 2 – €STR overnight index swap rate across maturities, provided by LSEG, as of June 30, 2023.

3.2 Dividend yield

To extract IVs from option prices, the dividend yield must be included as an input in the pricing formula (see equation (4) of Section 3.4). EMIR does not provide the dividend that will be paid by the underlying. Thus, rather than considering the expected dividend yields available from market data providers, we estimate the implied dividend yield directly from the data. Therefore, following Liu et al. [2021] we use the following equation

$$d_t(K,\tau) = -\frac{1}{\tau}\ln\left(\frac{V_t^c - V_t^p + Ke^{-r\tau}}{S_t}\right) \tag{1}$$

to estimate the implied dividend yield for each strike price K and maturity τ, where S_t is the underlying spot price at time t, V^c and V^p, are the values (i.e. the prices or premiums) at time t of call and put options, respectively, with strike price K and maturity τ reported in EMIR.

It is important to note that equation (1) provides an exact value for dividends computed from European options, but it is an approximation for American options. EMIR provides values for both calls and puts across the same strike prices and maturities, therefore we can apply equation (1) and compute the implied dividend yields. When one of the price values V^c or V^p is missing, we replace the not computed dividend with the median value computed across the same maturity. A potential data noise arises when there are contracts that share identical characteristics except for their option prices. From equation (1), it follows that any differences in input values will result in corresponding differences in dividend yields. To reduce this data noise, we cluster the contracts based on key features: maturity, strike, type, and style. For each cluster, we calculate the median option prices for call and put options and estimate the implied dividend.

Looking at the histograms of the computed dividend yield in Figure 3, it is possible to observe some outliers (e.g. some implied dividend yields are negative) and to assess the goodness of our approach. The x-axis displays the computed dividend yields, which vary for each underlying asset and also reflect the number of clusters. Our framework accounts for distinct expectations that investors may

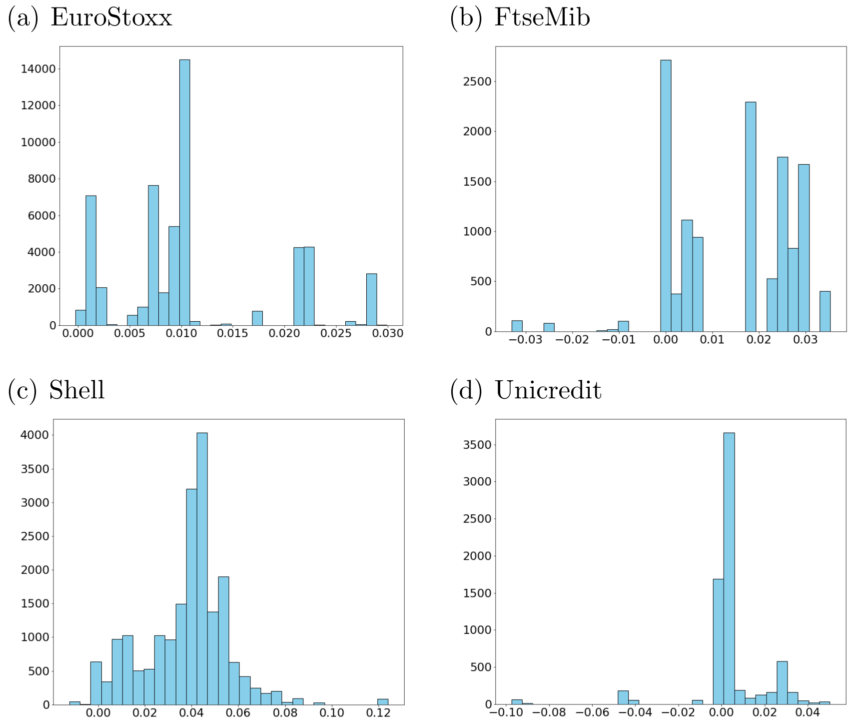


Figure 3 – Estimates based on EMIR data available at the Banca d'Italia. The histogram of estimated dividend yield as of June 30, 2023 for EuroStoxx, FtseMib, Shell and Unicredit.

have regarding the underlying asset, allowing for a realistic representation of market dynamics. In contrast, estimating dividend yields on a contract basis would result in a multitude of values on the x-axis, suggesting a wide range of investor expectations that would not align with observed market behavior.

As an alternative, for each underlying, we could estimate the implied dividend together with the volatilities by a single optimization algorithm, resulting in a single implied dividend for all maturities and strikes. This assumption implies that all investors expect the same dividend yield for all maturities, which may not accurately reflect market dynamics. Furthermore, implementing this solution would increase the algorithm processing time. Therefore, given the huge amount of data in EMIR and the need to minimize the computing time, we opted for the approach described above.

3.3 Additional data-tailored filter

Processing a huge amount of data can be challenging since data quality issues often occur. We face these issues in the preprocessing phase and during our initial examination of the equity option dataset. The initial dataset is obtained by following the cleaning and filtering steps described in Bianchi et al. [2025]. Then, we implement additional data quality procedures, resulting in a filtering process to address equity option specific issues. The EMIR equity options tailor made filter is four-step based and each step refers to a characteristic of the contract. The considered characteristics are option price, strike price, maturity and style of the

option-tailored filter	EuroStoxx	FtseMib	Shell	Unicredit
initial number of dataset observations	54,342	13,191	19,997	7,176
discarded maturities due to reporting errors	9	0	0	0
discarded strike price due to lower bounds	36	1	0	0
discarded strike price due to upper bounds	23	0	19	0
discarded options due to option price errors	9	239	8	60
number of options neither 'E', nor 'A'	544	0	0	0
final number of dataset observations	53,720	12,951	19,970	7,142

Table 3 – Based on EMIR data available at the Banca d'Italia. Discarded options in each step of the option-tailored filter process.

option. Therefore, at the beginning we discard options due to EMIR reporting errors about maturities (e.g. NaN). Then, as the second step, we define lower and upper bounds[6] for the strike price and discard options with strikes outside these bounds. In the third step we remove options with price below 0.001. Lastly, only European and American options are considered and we have the final number of options for which the IV will be estimated. For instance, in the EuroStoxx case just 1.12% of EMIR data are not considered and 98.88% of the dataset is used to compute the IVs (see Table 3).

3.4 Pricing engine

Estimating IVs from market data, such as EMIR equity options, involves solving a non-linear constrained optimization problem. The non-linearity arises because the option price (e.g., the Black-Scholes formula) does not have a linear relationship with the main input variable (i.e. the IV). Consequently, solving for IV requires algorithms capable of dealing with this non-linearity. Furthermore, the problem is constrained because volatility, by definition, cannot be negative. This imposes a lower bound constraint to ensure that the estimated IV is always positive. Additionally, an upper bound on volatility is assumed to avoid unrealistic estimates.

To estimate IVs we need an optimization algorithm that can handle both the non-linearity and the constraints of the problem. Since IV estimation is performed at the individual contract level, it is essential that the chosen optimization algorithm can find a solution in a reasonable computing time, especially given the potentially large number of contracts.

The algorithm discussed in this study enables us to extract IVs from European and American option prices by solving a minimization problem where the objective function is the relative difference between theoretical and observed option prices. The IV is the value that, when inserted in the pricing model, makes the theoretical price equal to the observed one. Therefore, determining IV through an optimization algorithm means finding a volatility value that minimizes the objective function, ideally reducing it to zero. Thus, volatility is the parameter of the algorithm and cannot be negative. We set lower and upper boundaries for this parameter at 2.5% and 300%, respectively. Both derivatives-based and derivative-free algorithms can iterate within these boundaries to find a solution to the minimization problem.

6 The upper bound for the strike price is 10 times the underlying price value and the lower bound for the strike price is 1e-6.

Given the massive amount of EMIR equity option data, speed becomes a crucial factor in determining the method to compute IV through the minimization of the objective function. As a result, a derivatives-free algorithm, such as the Brent method (see Brent [1971] and Brent [1973]), has been chosen for its efficiency and robustness. Derivative-free optimization algorithms represent a class of optimization techniques that do not depend on evaluating the derivatives of the objective function. Instead, these methods rely on direct evaluation of the function at various points to guide the search toward the minimum. The main strengths of derivative-free algorithms are robustness and efficiency. They provide accurate solutions in a reasonable time and, for this reason, they are particularly well-suited when a large number of optimization problems have to be solved, such as in the EMIR case.

As mentioned above, we rely on the optimization algorithm proposed by Brent [1971] to find the minimum of the objective (and continuous) function. This method is known for its robustness, rapid convergence and efficiency in terms of the number of function evaluations.[7] These features make it a popular choice for solving optimization problems and we select the Brent method as it is the fastest among the alternatives, such as the Newton-Raphson algorithm or other minimization methods available in the Python libraries (e.g. in *SciPy*).

The objective function is

$$f(\sigma) = \frac{P_{\text{EMIR}} - P_{\text{model}}(\sigma)}{P_{\text{EMIR}}}, \tag{2}$$

where P_{EMIR} is the option price observed in EMIR and P_{model} is computed through a model. Equation (2) represents a percentage distance, where the value of the objective function is expressed in percent of the observed EMIR price (P_{EMIR}). The P_{model} price is computed differently for European and American options because of the well-known difference given by the early exercise premium incorporated in the American option prices, which makes it not possible to assume a simple Black-Scholes-Merton (BSM) pricing formula for American options.[8]

Under the BSM model, the underlying price dynamics follow a geometric Brow-

[7] The algorithm begins by choosing three initial points a, b and c in the interval of interest. The points a and c satisfy the conditions $f(b)f(c) \leq 0$, $|f(b)| \leq |f(c)|$ and a may coincide with c. This initial setup ensures that the interval contains a point at which the function has a minimum. Next, inverse quadratic interpolation is applied to determine a point x at which a parabola passing through the points a, b and c reaches its minimum. If the calculated point x lies outside the interval $[a, b]$, the algorithm resorts to the bisection method. This consists of replacing x with the midpoint of the interval that has opposite sign to the function in x. For example, if $f(b)$ and $f(x)$ have the same sign, we replace x with the midpoint between b and c. After each iteration, the intervals are updated to maintain at least one point with a different sign than the other two. This process of updating the intervals ensures that the algorithm quickly converges to the minimum of the function. The algorithm continues to perform iterations of quadratic interpolation and bisection until the interval becomes sufficiently small or the value of the function at the point x is sufficiently close to zero.

[8] While the implementation of the BSM pricing formula is straightforward, we consider the Barone-Adesi and Whaley [1987] model implemented in the *finacepy* library to price American options.

	EuroStoxx	FtseMib	Shell	Unicredit
initial number of observations to be computed	53,720	12,951	19,970	7,116
IVs computed by BSM	53,680	12,949	303	174
IVs computed by BAW	40	2	19,667	6,942
IVs not computed	1,008	139	881	1,009
number of computed volatilities	52,712	12,812	19,089	6,107

Table 4 – Based on EMIR data available at the Banca d'Italia. Details on employed models to determine IV from each option. The not computed IVs does not show details about the employed model that could not determine the IV.

nian motion, that is

$$dS_t = (r - d_t)S_t\, dt + \sigma S_t\, dW_t^Q, \tag{3}$$

where S is the underlying spot price, r is the risk-free rate, d is the dividend yield, σ is the volatility and W_t^Q is a Brownian motion under the risk-neutral measure Q. This model provides an analytical solution for European options and the price (or premium) of a call option can be computed with the formula

$$C_t = S_t e^{-d_t \tau} \varphi(d_1) - K e^{-r\tau} \varphi(d_2), \tag{4}$$

where φ is the cumulative distribution function of a standard normal random variable with zero mean and volatility one, and d_1 and d_2 are

$$d_1 = \frac{\ln\left(\frac{S_t}{K}\right) + \left(r - d_t + \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}},$$
$$d_2 = d_1 - \sigma\sqrt{\tau}.$$

A similar formula allows us to compute the price of put options.

For American options, due to the early exercise premium, it is not possible to find a closed-form solution for the option. The theoretical prices can be found through the approximation proposed by Barone-Adesi and Whaley [1987] which takes this early exercise premium into account.

For American options, it may happen that the algorithm is not able to find a solution. While the Brent method is a robust and widely used algorithm for solving numerical optimization problems, it can present specific issues in the pricing of American options, particularly for negative discount rates (see De Donno et al. [2020] and Liu et al. [2021]).

In Table 4 we illustrate the results with respect to the employed models, which shows how our pricing engine worked for the chosen underlyings. Table 4 highlights that for EuroStoxx, starting from 53,720 reported option prices, we estimated 52,712 IVs. For FtseMib, we estimated 12,812 volatilities out of 12,951 reported option prices. For Unicredit, we estimated 6,107 IV out of 7,116 reported option prices, and for Shell, 19,089 out of 19,970. When the algorithm is not able to find a proper volatility, we observe that the lower and upper bounds of the value of such options are not satisfied (see Hull [2009]).

When the underlying is an equity index most of the reported options in EMIR are European, and when the underlying is an equity most of the options are American.

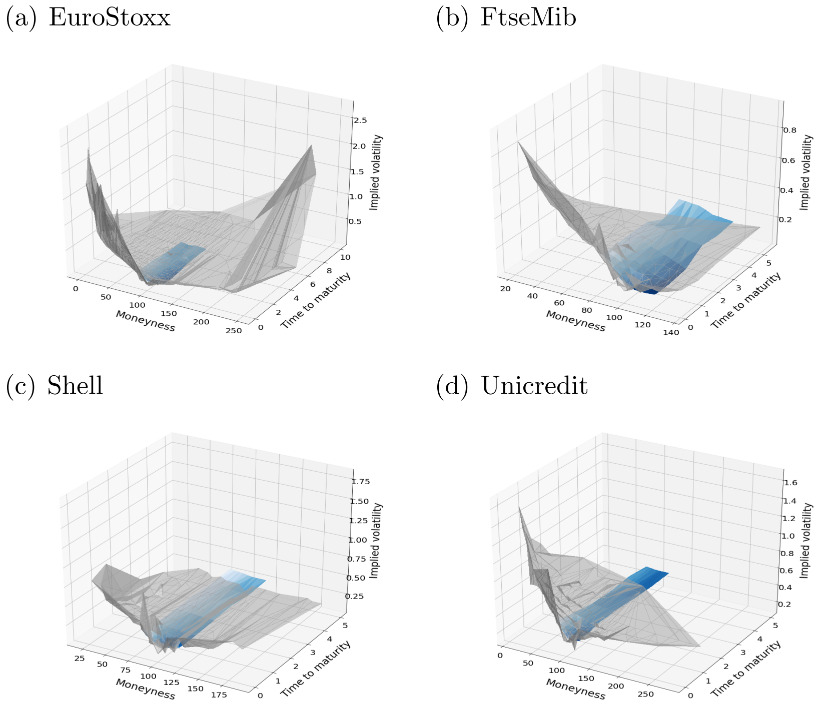


Figure 4 – Estimates based on EMIR data available at the Banca d'Italia. Comparison between EMIR IV surface (gray) and LSEG IV surface (blue). Ordered from top-left corner to bottom-right corner we have EuroStoxx, FtseMib, Shell and Unicredit.

4 Estimated implied volatilities

In this section we report the results of our analysis and we compare the IV surface extracted from EMIR prices to those obtained from a market data provider. Figure 4 represents a three-dimensional visualization of the results of our estimates and a comparison with data provided by LSEG on the same valuation date (June 30, 2023), for the four selected underlyings.

The gray surface represents the IV derived from EMIR equity options prices, whereas the blue surface represents the IV available from LSEG. We observe that the blue surface covers fewer moneyness values compared to the gray surface from EMIR. Specifically, the blue surface is constructed from approximately 90 data points, while the gray surface is computed from at least 5,000 data points. EMIR data provide broader coverage of moneyness values and, in some cases, cover a wider range of maturities than LSEG.

This comparison highlights the presence of outliers in the EMIR surface, which need to be flagged. We discuss a potential approach for flagging these outliers in the Section 5. It is important to notice that LSEG does not rely on actual contracts to extract the IV surface. Instead, it uses a model approximation and it is highly unlikely to find outliers in the data.

Table 5 shows the statistics of the differences between IVs extracted from EMIR and LSEG prices (i.e. $\sigma_{\text{model}} - \sigma_{\text{LSEG}}$). The lower number of observations in this table compared to the number of computed volatilities in Table 4 highlights the differences between the EMIR and LSEG data. The information content of EMIR

	EuroStoxx	FtseMib	Shell	Unicredit
n.obs	35,685	11,191	13,943	3,457
mean	-1.099	-0.289	-0.734	-0.649
std	3.856	2.014	5.046	2.444
min	-13.696	-13.307	-22.858	-20.176
25%	-2.618	-1.426	-1.510	-1.896
50%	-0.813	-0.370	-0.447	-0.367
75%	0.334	0.642	0.384	0.767
max	249.443	19.895	134.485	13.701

Table 5 – Estimates based on EMIR data available at the Banca d'Italia. We report the differences between IVs extracted from EMIR prices and those obtained from LSEG. Statistics are expressed in basis points.

data is richer compared to that of the data provider. This indicates that EMIR includes options not available in LSEG. The number of observations represents options found in both EMIR data and LSEG. By looking at the average and the median differences, we observe that estimated IVs are consistently smaller than those reported in LSEG. This difference is around or smaller than 1%.

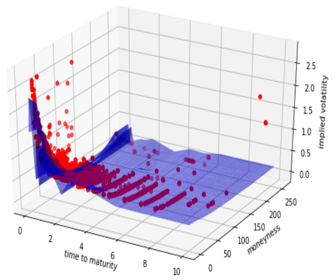
5 Quality checks

Now we propose a possible methodology to detect outliers in the IVs extracted from EMIR. This approach can be used for additional data quality checks applicable to all options that have underlying information available in EMIR. Outliers in the dataset can result from errors in reporting information related to options but can also be explained by the manual mapping step described in Section 2.

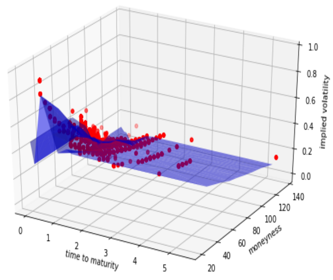
To address this issue, we approximate the IV surface estimated in Section 4 with a spline function, i.e. with a piecewise polynomial function with a specified degree of continuity imposed on its derivatives. We use these functions to achieve a smooth approximation of the data points represented by the estimated IVs. This EMIR surface approximation helps us identify outliers in the data by recognizing reporting errors that lead to anomalous data points. This process allows us to obtain a smooth approximation for each pair (knot) of moneyness-maturity within the two-dimensional grid of moneyness and maturity. Based on this objective, we follow a two-stage approach, considering the two dimensions: moneyness and maturity. First, for the sample of observed maturities, we approximate along the moneyness dimension while keeping the maturity fixed. This allows us to approximate each smile separately. Second, we build the surface by merging the approximations of the first step (the smiles) along the maturity dimension. This enables us to approximate the entire surface while maintaining the structure of each individual smile.

In more detail, in the first step, we interpolate EMIR volatilities for each maturity using a cubic spline. We focus only on maturities with enough observations, ensuring that the selected maturities accurately reflect the data distribution. In the second step, as we now have a smile for each maturity, we approximate volatilities across the maturity dimension to obtain values for each moneyness. This allows us to have an approximation for each pair of moneyness and maturity within the grid. Since we only estimate volatilities for the sample of knots (specific pairs of

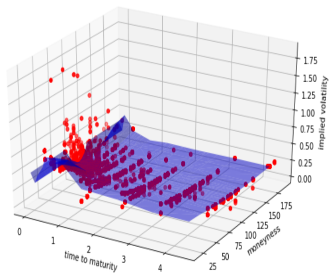
(a) EuroStoxx



(b) FtseMib



(c) Shell



(d) Unicredit

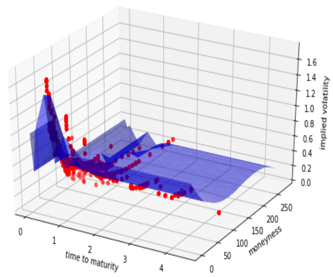


Figure 5 – Estimates based on EMIR data available at the Banca d'Italia. We show the results of the piecewise linear approximation for EuroStoxx (top-left corner), FtseMib (top-right corner), Shell (bottom-left corner) and Unicredit (bottom-right corner).

moneyness and maturity) defined through contracts reported in EMIR, we then approximate these volatility values across all possible pairs of moneyness and maturity in the grid, even though there may be some knots for which trades are not reported in EMIR. At this stage, we need to address another data quality issue. Indeed, as options move further into either out-of-the-money or in-the-money, and as maturities lengthen, the number of data points available for interpolation decreases. This scarcity of data makes it more challenging to construct a smooth surface defined at all knots, increasing the difficulty for the algorithm to identify and flag incorrect data. Therefore, we redefine the discretization of the moneyness array to obtain denser intervals around at-the-money values, where options are traded most, and wider intervals to collect more observations where options are out-of-the-money or in-the-money. This approach allows us to perform an approximation capturing more data points.

The approximation described in the second step is conducted in two different ways: (a) piecewise linear approximation and (b) cubic spline approximation, as in the first step. Figures 5 and 6 show the results for both methods: the red dots are the EMIR volatilities estimated in Section 4, and the blue surface is the approximation obtained by following our approach.

Figure 5 refers to (a) piecewise linear approximation, which is able to capture more of the behavior of volatility smile and sudden changes across various maturities. On the other hand, this method is sensitive to interval width and outliers, making it more challenging to distinguish outliers from good data. Results of the second method are provided in Figure 6. Here it can be seen that cubic spline

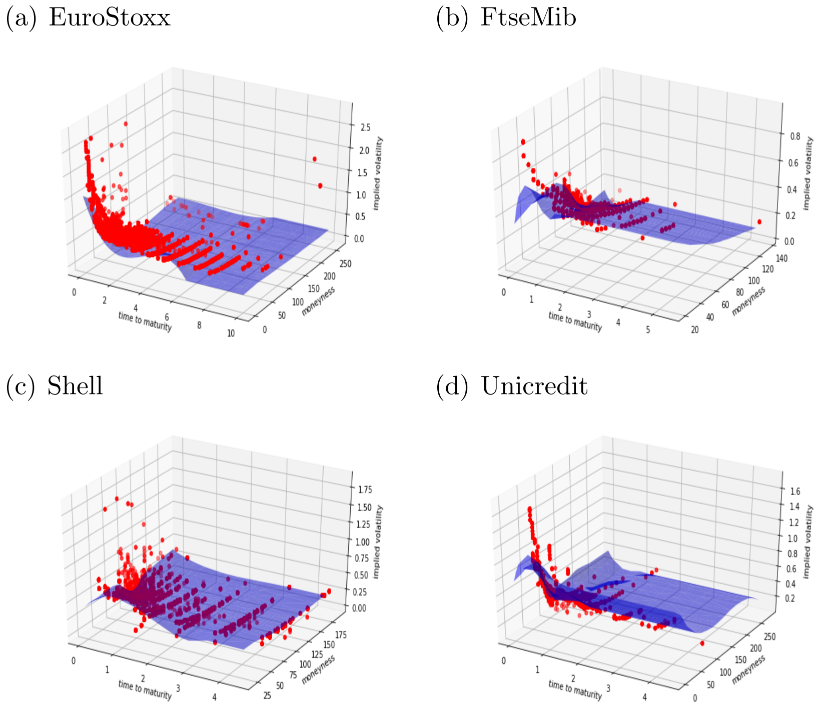


Figure 6 – Estimates based on EMIR data available at the Banca d'Italia. We show the results of the cubic spline approximation for EuroStoxx (top-left corner), FtseMib (top-right corner), Shell (bottom-left corner) and Unicredit (bottom-right corner).

provides smoother surfaces for EuroStoxx and Shell. On the other hand, for both FtseMib and Unicredit the shape of the surface is not a proper IV surface, at least for some knots of the grid. A reason for this pattern can be the number of options considered in the approximation. Indeed, either Unicredit and FtseMib has fewer data points than Shell and EuroStoxx, especially on the tails (wings) of the moneyness.

Table 6 compares the two methods, piecewise linear and cubic spline approximation, for the different underlyings, based on the difference $\sigma_{\text{model}} - \sigma_{\text{interpolated}}$. This difference can be used to flag outliers and perform advanced data quality analyses. We note that the piecewise linear method captures more variability and potential extremes, providing a more flexible but volatile fit (the surface is full of ups and

	piecewise linear approximation				cubic spline approximation			
	EuroStoxx	FtseMib	Shell	Unicredit	EuroStoxx	FtseMib	Shell	Unicredit
n.obs	51,491	12,761	19,020	5,942	51,491	12,761	19,020	5,942
mean	-0.127	-0.393	-0.188	0.909	0.167	0.183	-0.448	0.551
std	4.707	2.753	4.936	3.700	5.637	2.283	5.276	4.538
min	-12.754	-20.149	-19.733	-20.768	-15.885	-13.398	-30.812	-22.197
25%	-1.428	-1.254	-0.879	-0.367	-1.585	-0.928	-1.836	-1.926
50%	-0.228	-0.012	-0.129	0.581	-0.249	0.160	0.189	0.386
75%	0.687	0.993	1.060	2.148	0.879	1.107	1.079	2.158
max	248.609	21.411	138.612	38.042	249.142	19.181	141.020	40.944

Table 6 – Estimates based on EMIR data available at the Banca d'Italia. We report the differences between IVs extracted from EMIR prices and interpolated ones (both piecewise linear and cubic spline approximation). Statistics are expressed in basis points.

downs), while the cubic spline method tends to smooth out the data with extreme values (i.e. the surface is smooth). The choice between these methods should take into consideration the specific application and the importance of capturing variability versus having a smooth surface.

6 Conclusions and possible further analyses

IVs are so important that options are often quoted in terms of volatility rather than price. The purpose of this work is threefold. We define a framework (1) to estimate equity option IVs from EMIR option prices, (2) to compare these estimates with the data obtained from market data providers, and (3) to detect outliers by looking at the IV surface. Under this framework it is possible to conduct advanced data quality checks on EMIR option price data. The same approach can also be extended to other asset classes, for example options on interest rates (e.g., caps and floors). Additionally, the proposed approach can be easily parallelized on both the valuation date and the underlying dimensions in order to obtain the time series of estimated IV surfaces on the main underlyings.

EMIR data allow competent authorities to calculate implied volatilities based on actual transaction data, covering a broad spectrum of underlying assets and counterparties. This provides them with the ability to perform their own assessments without necessarily depending on third-party financial data providers. IVs are the best proxy for future volatilities. Having reliable IV surfaces based on real transactions helps setting up market monitoring tools to measure the risk-neutral asymmetry of investors towards rallies and sell-offs, or to calibrate complex asset pricing models useful to measure the probabilities assigned by market participants to future tail events related to a given underlying asset. IV is a standardized measure of risk in the same units as the asset returns and it can be used to compare the risk assessments of a given underlying asset by different market participants. Additionally, by tracking IV across different asset classes, such as equities, bonds, and currencies, it is possible to assess cross-asset risk and detect potential early warnings of market instability or financial crises.

As for possible monitoring tools, it is important to look at the IV smile, i.e. the pattern of IV across options with the same expiration date but with different strike prices, as well as at the term structure of volatility, i.e. the pattern of IV across options with the same strike price but different expiration dates. By looking at the smile, it is also possible to monitor the risk-neutral asymmetry, also referred to as tail risk (see Elyasiani et al. [2021]). Finally, reliable data of IVs in the more extreme wings, are very useful to calibrate heavy-tailed or stochastic volatility option pricing models, which allow one not only to perform advanced data quality checks but also to assess the probability of occurrence of tail events, under both the risk-neutral and the historical probability measure (see Bianchi and Tassinari [2020]).

Bibliography

AGOSTONI, G., A. IANIRO, A. JUKONIS, F. LENOCI, E. LETIZIA, AND G. SKRZYPCZYNSKI (2024): "Using trade-level derivatives data for macroprudential analysis," in *64th ISI World Statistics Congress*, https://isi-web.org/sites/default/files/2024-03/ottawa-2023_ips-1077-Grzegorz%20Skrzypczynski.pdf.

BARONE-ADESI, G. AND R. E. WHALEY (1987): "Efficient analytic approximation of American option values," *Journal of Finance*, 42, 301–320.

BIANCHI, M. L., B. SORVILLO, D. RUZZI, F. APICELLA, L. ABATE, AND L. DEL VECCHIO (2025): "EMIR data for financial stability analysis and research," in *IFC-Banca d'Italia Workshop on "Data science in central banking: enhancing the access to and sharing of data"*, Bank for International Settlements, no. 64 in IFC Bulletin, https://www.bis.org/ifc/publ/ifcb64_12.pdf.

BIANCHI, M. L. AND G. L. TASSINARI (2020): "Forward-looking portfolio selection with multivariate non-Gaussian models," *Quantitative Finance*, 20, 1645–1661.

BRENT, R. (1971): "An algorithm with guaranteed convergence for finding a zero of a function," *The Computer Journal*, 14, 422–425.

——— (1973): *Algorithms for minimization without derivatives*, Englewood Cliffs, NJ: Prentice Hall.

DE DONNO, M., Z. PALMOWSKI, AND J. TUMILEWICZ (2020): "Double continuation regions for American and swing options with negative discount rate in Lévy models," *Mathematical Finance*, 30, 196–227.

ELYASIANI, E., L. GAMBARELLI, AND S. MUZZIOLI (2021): "The skewness index: uncovering the relationship with volatility and market returns," *Applied Economics*, 53, 3619–3635.

HULL, J. C. (2009): *Options, futures and other derivatives*, Pearson, 7th ed.

LIU, S., A. LEITAO, A. BOROVYKH, AND C. W. OOSTERLEE (2021): "On a neural network to extract implied information from American options," *Applied Mathematical Finance*, 28, 449–475.

A holistic approach to EMIR data by the Central Bank of Ireland: Implementing the EMIR Roadmap

Gavin Beattie,[a,1] Angelos Karapappas,[a] Stamatoula Fragkou,[b]
Giuseppe Insalaco,[c] Ed Flood,[a] John Harte,[a] and Damien Quinn,[a]

[a]*Securities and Market Supervision Division, Central Bank of Ireland, Dublin, Ireland*
[b]*International and Finance Division, Central Bank of Ireland, Dublin, Ireland*
[c]*Markets Policy Division, Central Bank of Ireland, Dublin, Ireland*

Abstract. The European Market Infrastructure Regulation (EMIR) was introduced in response to the 2007 global financial crisis in order to address shortcomings in respect of the derivative market. Its focus is to increase transparency and reduce systemic counterparty and operational risk associated with derivatives and thereby help prevent future financial system crises. EMIR introduces the mandatory reporting of derivative trades entered into by EU counterparties. The data form a rich resource offering valuable insights to central banks and market authorities; however, users still face challenges due to the mixed data quality of the information reported, the complexity of the underlying derivative contracts, and the quickly evolving derivative product landscape.

The Central Bank of Ireland (CBI) holds a unique position in that we are responsible for safeguarding both monetary and financial stability. We regulate more than 10,000 firms across sectors through risk-based supervision, underpinned by a credible threat of enforcement. Our objective is to ensure financial stability, consumer protection and market integrity. To support this, we have a range of regulatory powers in the areas of authorisation, supervision and enforcement. Included in our wide mandate the CBI is the designated national competent authority for EMIR. Due to the open nature of the Irish economy, the presence in Ireland of many large multinational financial and non-financial corporations, and the fact that the

[1] The authors are grateful to staff of the Central Bank of Ireland for their helpful suggestions. This publication should not be reported as representing the views of the Central Bank of Ireland. The views expressed are those of the authors and do not necessarily reflect those of the Central Bank of Ireland.

Irish derivative market is very large compared to the Irish economy, it warrants increased supervisory attention.

As a result, we have developed the EMIR Roadmap, which is the key focus of this paper. The objective of the EMIR Roadmap is to build confidence in the quality of the EMIR dataset, resulting in increased access to EMIR to support supervision and the analysis of the Irish derivatives landscape. The EMIR Roadmap places in the hands of decision makers across the CBI high quality daily derivative information in a streamlined and accessible manner, influencing policy and supervisory decisions through empirical evidence. The EMIR Roadmap incorporates three pillars, each supported by a dedicated analytical supervisory dashboard:

1. monitoring and improving EMIR data quality (supported by the Rules Framework Dashboard);
2. empowering micro analysis for supervisory surveillance (supported by the EMIR Counterparty Report Dashboard);
3. empowering macro analysis for financial stability policy purposes (supported by the Network Analysis Dashboard).

Key words: EMIR, EMIR data quality, Rules Framework Dashboard, EMIR Counterparty Report Dashboard, network analysis.

1 Introduction

As set out in the strategy of the Central Bank of Ireland (CBI) for 2022-26, we aim to transform how we use data and analytics to drive our effectiveness as an intelligence-led organisation. As part of this initiative, the CBI is currently progressing a data strategy with the key objective of *'becoming deeply data driven'*.[2] Within that focus, EMIR is included as a previously relatively unexplored data resource that can provide valuable outputs.

The EMIR reporting obligation entered into force on 12 February 2014 and, since inception, it is noted that it has posed considerable challenges for national competent authorities (NCAs) leveraging the data to assist in fulfilling their mandates (see for example Rousová et al. [2015]). While the data quality issues are very heterogeneous, categories such as the uneven quality of reporting and the inclusion of abnormal values persist (for a full list of issues please see Agostoni et al. [2023] and Agostoni et al. [2024]). Irrespective of recognised data quality issues, after 10 years of derivative reporting, EMIR is now a well-documented, defined data set that can provide key insights in the pursuit of financial stability. The CBI recognises the value and importance of leveraging this key data resource and is in the process of delivering a strategy through the EMIR Roadmap with scope to foster supervisory engagement and promote the stability of the financial system.

[2] For further details please see "Our Strategy", https://www.centralbank.ie/publication/corporate-reports/strategic-plan and Central Bank of Ireland [2024].

In line with the CBI's focus on data driven approaches, the EMIR Roadmap ensures that supervisors, market analysts and policy makers have access to EMIR data, aiming at the highest level of data quality, such that it can be trusted and (re)used for multiple purposes.

Despite Ireland being categorised as a "small open economy", its links with domestic and international markets along with the relative size of its derivative markets signal its importance as a key player worldwide as a financial destination. Ireland plays a key role in derivative trading, due to the presence of multiple large international institutions, including banks and investment firms operating in this jurisdiction. In addition, Ireland is one of the largest fund domiciles in the world, the majority of which trade derivatives. To put this into context, Ireland has over 5000 Irish reporting counterparties under EMIR with a total outstanding notional of over €20tn, 37 times the country's GDP, and net valuation of €35bn in outstanding derivatives. Furthermore, there are 1.8 million outstanding derivative trades reported daily, placing Ireland as the 4th largest EU jurisdiction by number of trades and the 5th largest EU jurisdiction by outstanding notional.

This paper provides insight into the development and implementation of the EMIR Roadmap and its contributions in the area of supervision and financial stability. The EMIR Roadmap aims to address key challenges of supervisors, market analysts and policy makers by providing them with access to a valuable trustworthy data resource. The EMIR Roadmap also contributes from a technical standpoint addressing common issues faced by policy makers and supervisors in the analysis of a complex dataset. More importantly, the EMIR Roadmap is a holistic approach at resolving the issue of usability of data under imperfect data quality conditions to drive constant improvement.

2 The EMIR Roadmap

The EMIR Roadmap is a multi-faceted strategy implemented by the CBI. At its core the primary focus is a) improving EMIR data quality, thereby promoting accessibility to and usability of EMIR data, through supporting b) micro-financial analysis and c) macro-financial analysis.

The EMIR Roadmap incorporates three main pillars as follow (see Figure 1).

- **EMIR data quality.** We have implemented innovative supervisory approaches to improve EMIR data quality, thereby increasing people's trust in the quality of information. The starting point was to acknowledge that until experts can trust the data they would not use it. The intuition was that experts do not need data to be 100% accurate to be usable but they require confidence that holistically the data is of good quality. To achieve this we developed the Rules Framework Dashboard to monitor EMIR data quality and the risk-based approach to EMIR data quality to improve the data quality of the EMIR data set. This approach provides evidence-based assessments of EMIR data quality, thereby increasing confidence in the EMIR dataset as data quality improves. The objective is to provide users with metrics indicating the reliability of the data, and with evidence on whether the data quality is improving or not over time.

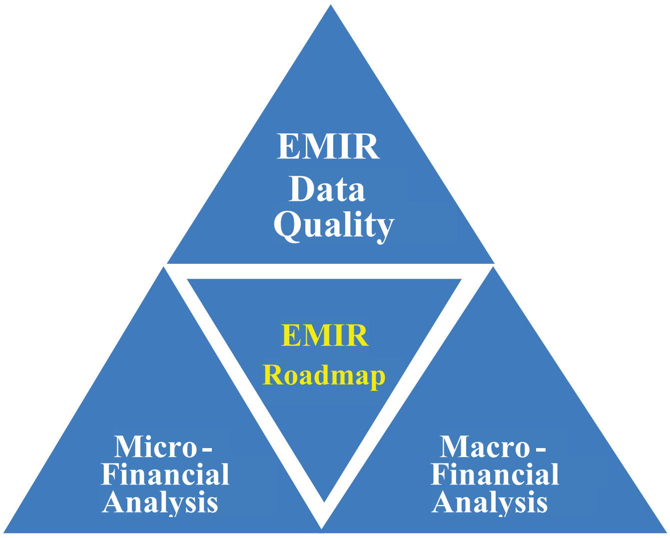


Figure 1 – Summary of the CBI EMIR Roadmap strategy.

- **Micro-financial analysis of EMIR supervisory engagement.** We developed the EMIR Counterparty Summary Report, which focuses on micro-financial analysis providing supervisors with the key oversight of the derivatives trading activities of their regulated firms. This dashboard tracks daily counterparties' risk profiles through EMIR trade data, displayed by a series of informative visuals, which indicate trends, identify potential vulnerabilities and map exposures.

- **Macro-financial analysis of EMIR stability of the financial system.** For users to have a complete view of derivatives risk they need to be able to assess the holistic derivative landscape. Users need to understand who are the primary central counterparties, to what sectors/users, in what proportions and through what channels clearing is being accessed. This is important for understanding the flows in a network of derivative relationships transferring risks across the financial system. For this purpose, we have under development a computational network approach to display interconnectedness of the clearing landscape (i.e. interactive maps of counterparties, clearing members and central counterparties connected through clearing channels), assisting in the identification of potential contagion and risk vulnerabilities. The EMIR Network Analysis Dashboard visuals provide supervisors with access to an interactive tool to further drill down and assess the interconnectedness of the derivatives market. Each of the above tools supports users across the CBI by providing key insights into firms' derivatives trading activity and discernible network analysis of the interconnectedness of the Irish derivative landscape.

Figure 2 – Rules Framework Process for EMIR.

2.1 EMIR Data Quality

Rules Framework Dashboard

We developed the Rules Framework in 2022. The Rules Framework is a centralised, metadata-driven engine whose purpose is to measure the conformance of data to predefined rules. Additionally, a dashboard has been developed that allows the performance of these rules to be tracked on a predefined schedule. The Rules Framework allows users to author their own rules (scripts) in SQL, Python or R and to check compliance of individual records to a requirement (e.g. a regulatory requirement or a data quality requirement). The framework hosts the rules, executes them on a schedule and stores the results in a relational database. Users access performance data on rule results through a Power BI dashboard. To date, the main use of the Rules Framework has been to check compliance with EMIR and SFTR data quality rules, and it is being expanded to check collective investment schemes compliance with regulation.

In respect of EMIR, the Rules Framework hosts each of the EMIR data quality rules as agreed with ESMA [2023] and ESMA [2024].

These rules run on a daily basis and provide two measures.

- The performance and quality of the overall EMIR dataset and for each specific data quality rule. The Rules Framework has the functionality to measure data quality over time. This allows the CBI to identify if improvements in data quality have been observed at an overall EMIR dataset level and in respect of each data quality rule over a specific period. It also provides the evidence of how reliable the data are from an end user's standpoint.

- The worst performing counterparties under each of the data quality tests. The Rules Framework provides the name of the top ten worst performing counterparties on an absolute basis for each of the specific data quality tests. This supports supervisors to implement the risk-based approach to EMIR data quality.

Risk-based approach to EMIR data quality

With the assistance of the Rules Framework, the CBI has initiated a new risk based approach to EMIR data quality. This approach involves providing significant offending counterparties with the results of the relevant EMIR data quality tests in which they have underperformed. The counterparties also receive details on how the data quality tests are performed. The counterparty is required to perform specific data quality test(s) on a quarterly basis and provide the results to the CBI. If no improvement is identified, the counterparty must detail why this is the case and provide a timeline for improvement. The CBI advises that the improvement in the quality of EMIR data is the responsibility of the board of directors of the relevant counterparty. This approach has evidenced improvement in the majority of EMIR data quality tests performed by the CBI. The CBI will also prioritise engaging in respect of EMIR data quality issues that affect the analysis of the EMIR data. Where CBI supervisors, market analysts and/or policymakers identify data quality issues that impact their analysis of derivative trading, the CBI will engage directly with the relevant counterparty in order to ensure prompt remediation of the issue. The CBI, generally, does not clean the EMIR data prior to use. If there are data quality issues impacting analysis of the data, the first step is to try to remediate these issues in order to improve the quality of the EMIR data and enable accurate analysis of them. Evidencing improvement in data quality and providing a measure for the overall EMIR data quality are vitally important in developing confidence in the EMIR data and thereby increasing use and analysis of it across the CBI.

2.2 Micro-financial analysis of EMIR

EMIR Counterparty Summary Report

The key objective of the EMIR Counterparty Summary Report (CP report) is to put the derivative exposures of supervised firms into the hands of supervisors, so that they can understand and interrogate firms on it. Previously, when supervisors requested EMIR data on a supervised firm, this was provided through an excel spreadsheet that listed all the derivative trades and associated key fields of the firm for a specified date. In order for supervisors to gain any meaningful insight from these data, the supervisors would have to perform analysis of the data (e.g., aggregate it and forex conversions) and have some insight into EMIR reporting to understand what each of the key fields meant. The EMIR CP report provides supervisors with access to regulated firms' derivative exposures through an interactive dashboard.

The EMIR CP report provides insight to supervisors on the derivative activity and associated risk profile of individual reporting counterparties. The information collected regards every single derivative trade state report ranging from numerical information (notional amounts, collateral posted and variation margins) all the way to metadata on the profiles of the parties completing the transactions (type of entity, unique identification etc.). By using the information from the trade state report, we resolve the issue of missing intraday input. The processing is conducted in SQL and the data are not truncated or in any other way condensed. Rather, we

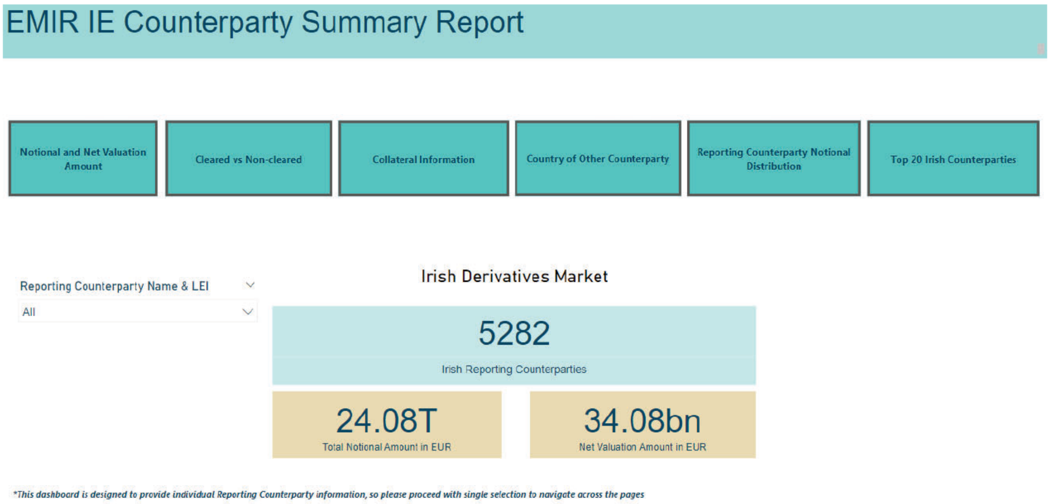


Figure 3 – Summary of information included in the EMIR CP Report.

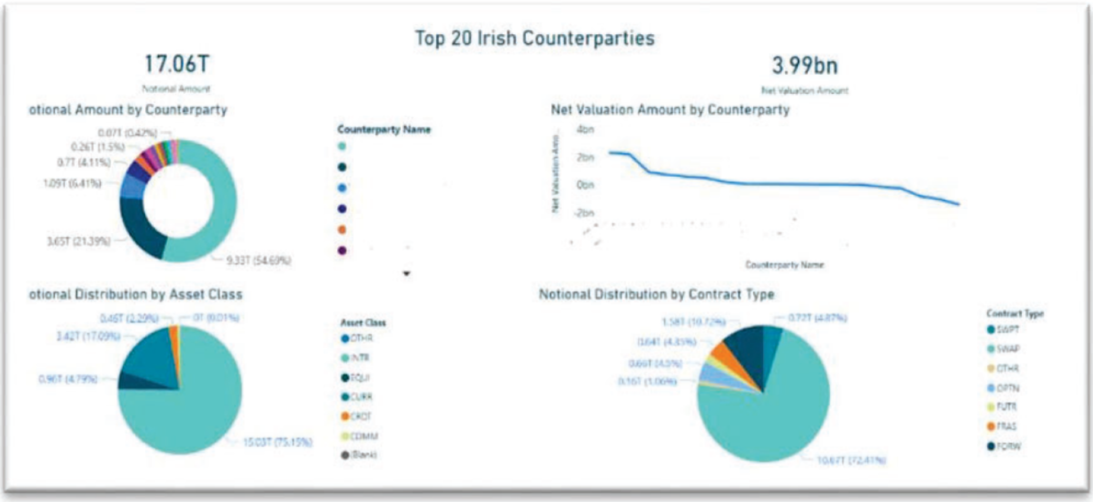


Figure 4 – Redacted TOP 20 Irish counterparties by notional amount in euro from EMIR CP Report.

retain all information. Filtering and drill down options are available for in depth analysis and summarisation of key information. Specifically, the EMIR CP Report (see Figure 3 below for summary) provides analysis on all Irish counterparties by notional amount in euro on a daily basis (Figure 4). Additional information provided in the report includes:

- clearing activity, including the number of cleared transactions and through where they are cleared (Figure 5);
- exchange of margin including initial margin and variation margin received and posted (Figure 6);
- distribution of derivatives, across asset class and contract type;

Figure 5 – Redacted individual counterparties cleared vs non-cleared trades and notional amounts by central clearing counterparties (CCP's) in euro from EMIR CP Report.

Figure 6 – Redacted individual counterparties initial and variation margins received and posted distribution.

- derivative exposures, including to what other counterparties and to what geographic jurisdictions most exposed.

2.3 Macro-financial analysis of EMIR

EMIR network analysis design

The key objective of the EMIR network analysis is to display the holistic Irish derivative landscape, so that market analysts and policymakers can analyse it to identify potential vulnerabilities and inform policymaking. Derivatives and their enabling infrastructure of central counterparties, clearing members, collateral and liquidity providers are increasingly pivotal in explaining the behaviour of the markets and driving policy response. Many of the recent financial crises had derivative use as a contributor (e.g. Lehman, AIG, Archegos, and LDI funds).

Figure 7 – Example of network for cleared trades, by notional amount in euro (averages) and by count of trades, filtered for interest rate as asset class and swaps as contract type where counterparties and other counterparties are separated by colour and size of bubbles represents notional amount in euro.

The EMIR Network Analysis Dashboard maps the interconnectedness of the Irish derivative trading and clearing landscape and provides an interactive dashboard and visuals to display the results. The dashboard is currently under development with the first iteration delivered, which is to map the clearing channels of the Irish derivative landscape, i.e. central counterparties and clearing members through to counterparties. Future iterations will map margin flows, bilateral derivatives trading and different types of contracts. To perform network analysis, we initially classified cleared and non-cleared transactions at counterparty pair level. In the next steps of our analysis, we intend to use the notional as the weighted node and use eigenvector centrality to measure the transitive influence of nodes (work in progress).

We map this through visuals for all asset classes, contract types and entities to assist the identification of potential sources of systemic and concentration risk. To support this we have built a report through Power BI which contains information on central counterparties, clearing members and reporting counterparties by taking into account the notional, valuation amounts, number of transactions and clearing members information (see for example Figure 7). We have significant interest in clearing members and the concentration of cleared derivatives. Contract type where counterparties and opposing counterparties are separated by colour and size of bubbles represents notional amount in euro. In addition, this approach allows us to take into account additional information such as the number of trades between counterparties along with the identities of the clearing members and central counterparties.

This process allows us to break down the network into manageable parts. For example, when focusing on interest rate asset classes for the same date (Figure 7), network relationships start to emerge and key players for cleared trades reveal their interconnectedness and interdependence. The next steps will focus on including more than one-to-one relationships, separating types of entities by colour, sizing the nodes based on notional and valuation amounts, and developing overnight metrics for the networks.

3 Outcomes

3.1 Improvement in EMIR data quality

The implementation of the EMIR Roadmap has resulted in an improvement in the majority of EMIR data quality tests performed (Fgiure 8). All of the indicators are in line with ESMA guidelines and reflect the improvement of Irish EMIR data as measured at a European level (ESMA [2023]).

Name of Data Quality Test
Missing Variation Margin
Outdated Valuation
Outstanding Transactions
Inconsistent Corporate Sector
Missing Valuations
Client Codes
Inconsistent Financial Sector
Matching
Pairing
Outstanding Positions
Blank Collateralisation
Inconsistent Counterparty Nature
Blank or Abnormal Maturity
Default Maturity
Duplicate Trades

Figure 8 – Evidence of improvement across a range of data quality indicators from the Rules Framework 2022-2023. Those in green demonstrated improvement.

EMIR enforcement case

On 28 November 2023, the CBI reprimanded and fined an investment fund €192,500 pursuant to the European Union (European Markets Infrastructure) Regulation 2014, as amended (the EMIR Regulations), for breach of its reporting obligation under Article 9(1) of the European Markets Infrastructure Regulation (EMIR) which requires details of all derivative contracts to be reported to a registered trade repository no later than the working day following the conclusion of the contract. This enforcement case demonstrated to the industry[3] the importance of ensuring complete, accurate and timely regulatory reporting and the consequences of failures concerning regulatory reporting (for further details please see Central Bank of Ireland [2023]).

The accompanying public statement included the following viewpoint from the CBI's Director of Enforcement and Anti-Money Laundering.

> *This case highlights the importance of timely and accurate data reporting. The reporting obligations under EMIR and other sectoral legislation increase transparency and enable the CBI to obtain a complete picture of each firm's operations, to fully understand the risks facing firms operating in securities markets, and thereby to address systemic risk. Incomplete or inaccurate data actively hinders market monitoring processes and activities.*
>
> *Firms must have appropriate oversight of data reporting from Board level down, including where data reporting is delegated or outsourced. The delegation of reporting obligations must be appropriately managed in order to avoid confusion between the delegates as to their respective reporting responsibilities.*
>
> *We have reiterated the importance of data quality and of EMIR reporting to industry over a number of years and in each Securities Markets Risk Outlook Report we have published since 2021. The gap in data reporting which gave rise to this investigation only became apparent to the ICAV upon a review of its EMIR reporting arrangements, prompted by the CBI's letter to industry in 2019 on this theme.*
>
> *We also expect firms to bring material failures to our attention at the earliest opportunity and to act expediently to address identified issues. This investigation found that, despite the ICAV identifying in May 2020 that thousands of its derivative trades had not been reported to a trade repository in breach of its EMIR reporting obligation, the ICAV only notified the CBI of this failure following engagement initiated by the CBI.*
>
> *Compliance by the industry with data reporting obligations will continue to be an area of focus for the CBI.*

[3] Coverage of the decision https://www.rte.ie/news/business/2023/1130/1419344-central-bank-fines-globalreach-multi-strategy-icav/ and https://www.irishtimes.com/business/2023/11/30/central-bank-fines-fund-192500-for-failing-to-report-financial-contracts/.

The large number of Irish firms that performed reviews of their EMIR regulatory reporting and conveyed the outcomes to the CBI evidences the impact on industry of the EMIR enforcement case. These reviews resulted in further improvements in EMIR data quality.

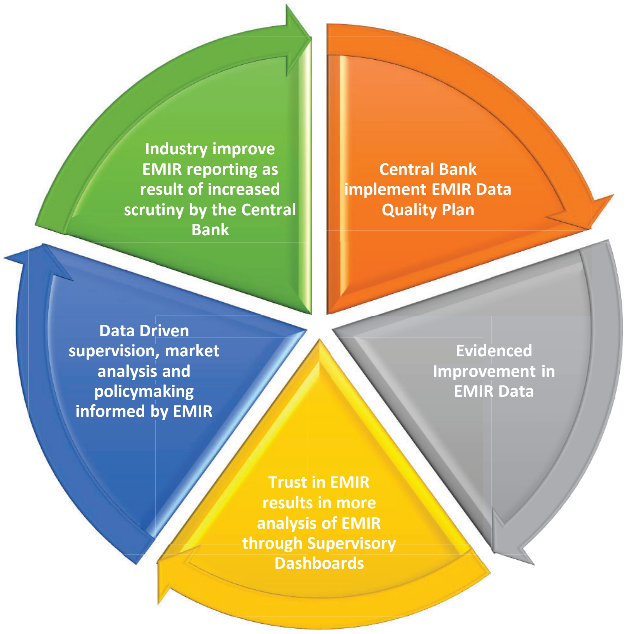


Figure 9 – Life cycle of improvement in usage of EMIR through the EMIR Roadmap across the CBI and the financial industry improving analysis and policymaking decisions.

Increased access to EMIR

An increased number of supervisory and market analysis teams across the CBI now have greater access to EMIR data, either directly through the underlying EMIR database or indirectly through analytical dashboards (as described above), than at any other previous time since EMIR reporting went live in 2014. This dataset is being leveraged more widely across the CBI to inform our supervisory approach.

4 Conclusions

The CBI is focussed on delivering a data driven strategy to ensure data can be trusted, reused and analysed in order to assist in the delivery of our wide mandate. The EMIR Roadmap is an integral part of this forward-looking strategy. The vision for the EMIR Roadmap is that supervisors, market analysts and policy makers are provided with evidenced high quality EMIR data that they can trust. Once trust is built, data are provided to supervisors, market analysts and policymakers through interactive dashboards, which encourages analysis to inform supervision, market analyses and policymaking. The more EMIR data are accessed and analysed around the CBI, the more this indicates to industry the importance of ensuring complete, accurate and timely EMIR reporting, again improving standards.

Achieving the above does not require perfectly cleaned data, nor expensive data cleansing. Instead, we rely on increased access and analysis of EMIR data to inform where large data quality issues are and direct engagement with counterparties for them to remediate. Counterparties understand that the EMIR data are part of our market and supervisory analysis and that there will be consequences for submitting poor quality EMIR data and, as a result, improve the quality of the EMIR data being reported. All of the above leads to building on trust between users and the firms engaging in derivative trading.

Bibliography

G. Agostoni, L. De Charsonville, M. D'Errico, C. Leonte, and G. Skrzypczynski. Anomaly intersection: disentangling data quality and financial stability developments in a scalable way. In *IFC-Bank of Italy Workshop on "Data Science in Central Banking: Applications and tools"*, number No 59 in IFC Bulletin. Bank for International Settlements, 2023. https://www.bis.org/ifc/publ/ifcb59_32_rh.pdf.

G. Agostoni, A. Ianiro, A. Jukonis, F. Lenoci, E. Letizia, and G. Skrzypczynski. Using trade-level derivatives data for macroprudential analysis. In *64th ISI World Statistics Congress*. 2024. https://isi-web.org/sites/default/files/2024-03/ottawa-2023_ips-1077-Grzegorz%20Skrzypczynski.pdf.

Central Bank of Ireland. Enforcement Action. *Press release*. https://www.centralbank.ie/news/article/press-release-globalreach-fined-192-500-and-reprimanded-by-central-bank-of-ireland-under-emir-30-november-2023, 2023.

Central Bank of Ireland. Fulfilling our public sector duty assessment & action plan. https://www.centralbank.ie/docs/default-source/publications/corporate-reports/strategic-plan/our-strategy/fulfilling-our-public-sector-duty-assessment-action-plan.pdf, 2024.

ESMA. Report on quality and use of data. https://www.esma.europa.eu/sites/default/files/2024-04/ESMA12-1209242288-852_2023_Report_on_Quality_and_Use_of_Data.pdf, 2023.

ESMA. Follow-up report to the peer review into supervisory actions aiming at enhancing the quality of data reported under EMIR. https://www.esma.europa.eu/sites/default/files/2024-04/ESMA42-2004696504-7771_Follow-up_Report_to_EMIR_data_quality_peer_review.pdf, 2024.

L. Rousová, M. Osiewicz, and G. Skrzypczynski. 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, 2015.

Leverage, liquidity and concentration: Vulnerabilities of GBP liability-driven investment (LDI) funds through the prism of regulatory data

Antoine Bouveret[a,1] and Hadrien Leclerc[a]

[a]*European Securities and Markets Authority, Paris, France*

Abstract. We assess risks related to funds pursuing liability-driven investment strategies by merging entity and activity-level regulatory datasets. The analysis points to high concentration risks stemming from large holdings of UK sovereign bonds and a high degree of portfolio overlap among funds. We also document the use of leverage sourced from interest rate derivatives and repo borrowing. We perform a liquidity stress test by estimating liquidity demands following a large interest rate shock. Funds would not have enough cash to meet margin calls and collateral demands but they could mobilise unpledged bonds to raise liquidity. Our work illustrates how liquidity stress testing can be performed using EMIR data.

Key words: liability-driven investment funds, EMIR, SFTR, stress test.

[1] The views expressed are those of the authors and do not necessarily reflect the views of the European Securities and Markets Authority (ESMA). Any error or omissions are the responsibility of the authors.

1 Introduction

In September 2022, United Kingdom (UK) gilt yields surged following the announcement of an expansionary budget by the UK Chancellor, including through large unfunded tax cuts (Pinter [2023]). The surge created acute liquidity issues for some investment funds domiciled in the European Union (EU) pursuing liability-driven investment (LDI) strategies on behalf of UK pension funds (hereafter EU GBP LDI funds). As yields rose, these funds faced liquidity demands stemming from margin calls on their interest rate derivatives (IRDs) exposures and collateral requests on repo operations (ESRB [2023]). LDI funds had to raise cash to meet those liquidity demands and started liquidating UK sovereign bonds, putting additional downward pressure on bond prices. Stress in the UK sovereign bond market prompted the Bank of England to intervene through asset purchases to mitigate the self-reinforcing feedback loop (see Chen and Kemp [2023]). Following the crisis, EU authorities implemented measures to mitigate leverage-related risks for EU GBP LDI funds (see ESMA [2024]).

The LDI event shows how the combination of risks related to leverage, liquidity and concentration can result in financial stability issues stemming from non-bank financial institutions.[2]

In this article, we merge different regulatory datasets (entity and activity-level data) to assess vulnerabilities related to EU GBP LDI funds.[3] LDI funds are highly exposed to concentration risk stemming from their exposure to interest rate risk related to UK yields. Those risks are magnified by the use of leverage through interest rate swaps and borrowing collateralised by UK bonds. We perform a liquidity stress test and find that most funds would not have enough cash to meet liquidity demands stemming from a large interest rate shock and would have to mobilise unpledged sovereign bonds.

2 A review of LDI business models

UK defined-benefit pension funds provide guaranteed returns to future pensioners. To do so, pension funds must invest in long-dated assets (to match the duration of their future payments to pensioners) and ensure that the returns on those assets are high enough to meet the guaranteed returns. In that context, some smaller UK pension schemes use pooled funds pursuing LDI strategies rather than investing directly into long-term assets, as pooled LDI funds offer a more cost-effective

2 The collapse of Archegos, a US family office with concentrated exposures on a few technology stocks, is another example of those vulnerabilities. In the case of Archegos, the entity used total return swaps on the underlying stocks, and mirrored its exposures across several prime brokers, to obtain synthetic leveraged exposures to a few listed firms. When the price of some stocks started declining, Archegos failed to meet margin calls, forcing its counterparties to liquidate their equity exposures (used to hedge the total return swap), resulting in sharp price declines for those stocks, and large losses for prime brokers (up to $10bn overall). For further details see Bouveret and Haferkorn [2023].

3 Those datasets relate to entity-level data collected under the Alternative Investment Fund Manager Directive (AIFMD) and activity-level data on derivatives from the European Market Infrastructure Regulation (EMIR) and on repo activity from the Securities Financing Transactions Regulation (SFTR).

hedging solution (Chen and Kemp [2023]). Those GBP LDI funds are typically domiciled in Ireland and Luxembourg, two global asset management centres.

Reducing the duration mismatch

Schematically, a defined benefit pension fund employing LDI strategies will invest in LDI funds to reduce the duration mismatch between its long-term guaranteed liabilities and its asset portfolio. The fund will target a long duration by investing directly in long-dated sovereign bonds (including inflation-linked bonds).[4] However, given the limited supply of long-term gilts, transaction costs and other factors, LDI funds will also increase their duration synthetically, by entering into long-term IRDs.[5] LDI funds are exposed to rising interest rates as they receive a fixed rate and pay a floating rate, thereby reducing the duration mismatch for the pension fund.

Reducing the return mismatch

Given the low returns on sovereign debt, LDI funds will use leverage to reduce the mismatch between the return on assets and the guaranteed returns promised to future pensioners. Funds can obtain leverage through repo borrowing, by pledging some of its long-dated bonds as collateral, and using the cash to purchase higher-yielding assets. Leverage can also be obtained using derivatives (synthetic leverage), as the LDI exposure to interest rates is a fraction of the initial margin posted.

Leverage in LDI funds

The use of leverage is visible in entity-level data (AIFMD), where funds report a range of information, and in activity-level data (SFTR for repo and EMIR for derivatives). We use a sample of 366 EU GBP LDI funds, identified using the name of the fund and information from supervisory authorities. Figure 1 (left panel) shows a decomposition of EU GBP LDI funds' leverage into repo borrowing and derivatives over 2021-2023. In mid-2022 (before the mini-budget crisis), the net asset value (NAV) of GBP LDI funds amounted to around €133bn, while their assets under management (AuM), which measures gross exposures, amounted to €673bn, resulting in a gross leverage of around 5 times the NAV. The net leverage of LDI funds, measured through the commitment method (which takes into account netting and hedging effects) reached 3.4 times the NAV over that period. Net leverage stemmed from repo borrowing for around €140bn and synthetic leverage through derivatives for €178bn.

Liquidity of LDI funds

The use of leverage exposes funds to default risk (as leverage magnifies profits and losses) and liquidity risks due to potential liquidity demands related to derivatives

[4] LDI funds invest in such bonds and enter into inflation swaps to hedge their portfolio against inflation risk, since guaranteed returns are also tied to inflation, see Barria and Pinter [2023].

[5] For further details on the use of swaps see Pinter [2023].

and repo borrowing. Liquid assets held by LDI funds are composed of UK sovereign bonds,[6] money market fund (MMF) shares and cash (Figure 1, right panel). Overall, liquid assets amounted to around 20% of AuM in June 2022 and were highly concentrated toward UK gilts (around 80% of total liquid assets).

(a) NAV, repo and OTC derivatives (b) Composition of unencumbered liquid assets

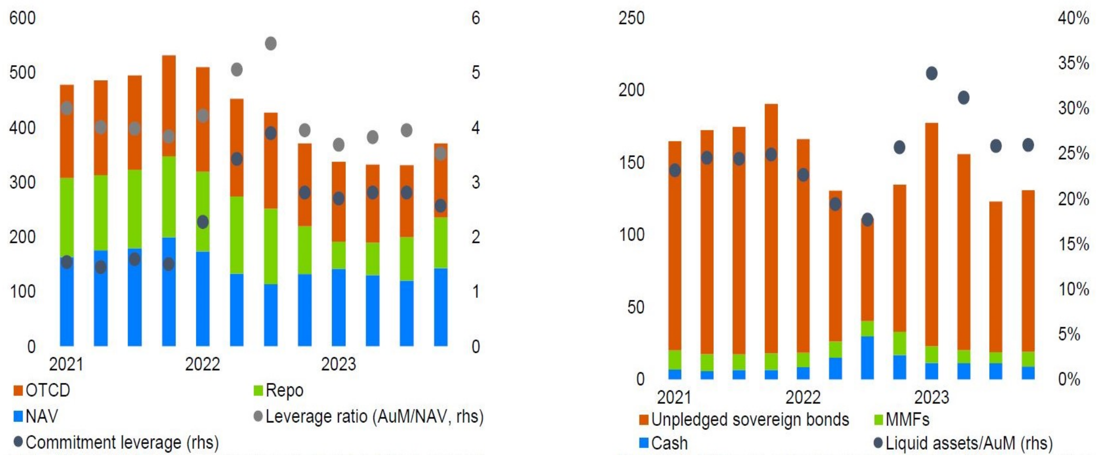


Figure 1 – LDI's assets and leverage. Panel (a) displays EU GBO LDI funds net asset value and borrowing through repo and borrowing embedded in OTC derivatives in €billion. Sources: AIFMD and ESMA. Panel (b) displays cash MMFs and unpledged sovereign bonds in €bn and ratio to AuM in %. Sources: AIFMD and ESMA.

3 Risks and vulnerabilities related to LDI funds

3.1 Analysis of potential risks

Concentration risk and interest rate exposure

LDI exposures are extremely concentrated towards UK long-dated gilts and the funds are highly exposed to interest rate risk. By design, LDI funds hold mainly long-term sovereign bonds, which implies that an increase in interest rates would result in large mark-to-market losses on bond holdings given the long duration of the instruments. In addition, the decline in the value of UK gilts posted as collateral would trigger sizeable collateral requests on repo borrowing, as any decline in the value of the collateral would trigger collateral calls from the LDI counterparty. An increase in rates would also result in liquidity demands for LDI funds through its effect on the valuation of interest rate derivative (IRD) positions. As rates increase, counterparties paying floating rates (like LDI funds) would face mark-to-market losses on their swaps and hence would have to post variation margins to their counterparties. The LDI sector itself is also highly concentrated: the 366 LDI funds in the sample are managed by only 7 managers.

6 AIFMD only provides high-level information on sovereign bond holdings. We estimate the amount of unpledged sovereign bonds by subtracting from AIFMD long sovereign exposures the amounts of GBP bonds pledged as collateral using data from SFTR.

Amplification factors

Interest rate risk is further amplified by the high market footprint of LDI funds in some segments of the UK sovereign bond market. First, LDI funds own a large share of long-dated and inflation-linked gilts, and the funds account for most of the trading activity in those segments (Barria and Pinter [2023]). Second, Pinter [2023] shows that LDI funds are by far the largest borrowers in the GBP repo market using UK nominal and inflation-linked bonds. In addition, LDI funds also have the largest net exposures in the inflation swap market. Moreover, LDI funds have a high degree of portfolio overlap: the specific bonds held by individual funds also tend to be held by other funds.

3.2 Empirical analysis

Figure 1 already shows the high degree of asset concentration. We extend the analysis by estimating portfolio overlap measures and focusing on counterparty concentration in derivatives markets, using end-2023 data.

Portfolio similarity among LDI funds

Since AIFMD does not provide granular portfolio information, we rely instead on SFTR data, where instrument-level information is available on bonds pledged as collateral by LDI funds. We measure the collateral overlap using cosine similarity (see Girardi et al. [2021]). The cosine similarity between the collateral portfolio of two LDI funds w_i and w_j, where w is the vector of portfolio weights for each fund, is defined as:

$$\text{Cosine Similarity}_{i,j} = \frac{w_i \cdot w_j}{\|w_i\| \|w_j\|}$$

where:

$$w_i \cdot w_j = \sum_{k=1}^{n} w_{i,k} w_{j,k}$$

and

$$\|w_i\| = \sqrt{\sum_{k=1}^{n} w_{i,k}^2}, \quad \|w_j\| = \sqrt{\sum_{k=1}^{n} w_{j,k}^2}$$

The similarity measure ranges between 0 (no overlap) to 1 (perfect overlap). Figure 2 indicates that collateral overlap is very high for some funds and instruments. Large edges (representing high cosine similarity values) between nodes (LDI funds) are visible, implying that most funds post the same gilts as collateral for repo borrowing. The combination of high market footprint and high portfolio overlap implies that if LDI funds were to liquidate sovereign bonds, the downward pressure on prices would be high (as LDI funds own a large share of the market) and would spread to other LDI funds (due to portfolio overlap).[7]

[7] The combination of high market footprint and high portfolio overlap is not unique to LDI. MMFs have similar vulnerabilities, see Bouveret and Danieli [2021] for further details.

Figure 2 – Collateral similarity for LDI funds. Only similarity measures above 0.5 are shown. Source: SFTR and authors' calculation.

Counterparty concentration in derivatives markets

We use EMIR data to visualise bilateral exposures between LDI funds and their counterparties. Given the large number of LDI funds, we aggregate the exposures by LDI managers to simplify the visualisation. Figure 3 shows the resulting network of derivatives exposures between LDI funds (aggregated at manager level) and their counterparties. First, LDI derivatives exposures are concentrated in three main managers (represented by the size of the blue triangles), while other managers have lower exposures. Second, most LDI counterparties are non-EU banks: mostly UK banks (red squares) and banks domiciled in other non-EU jurisdictions (orange squares). In contrast, EU banks (blue squares) play a marginal role, as indicated by the two small nodes. In addition, a few UK banks account for most of the derivatives exposures (three largest red squares) and tend to be exposed to the same LDI managers. In other words, LDI funds use several counterparties, which tend to be the same banks for most LDI funds.

4 The mini-budget stress of 2022

The risks and vulnerabilities analysed previously were put to the fore during the mini-budget crisis of September 2022.

Following the announcement of the mini-budget, UK yields spiked between 22 and 29 September: the 30-year rate jumped 130 basis points over that week, an unprecedented move (Figure 4, left panel). The surge in yields triggered a large drop in bond prices, including long-term bonds pledged as collateral; the value of some of these plunged by more than 40%. The fall in bond prices was even larger for inflation-linked bonds, which are mostly held by LDI funds (Pinter [2023]). Dunne et al. [2023] estimate that Irish-domiciled LDI funds experienced a 40%

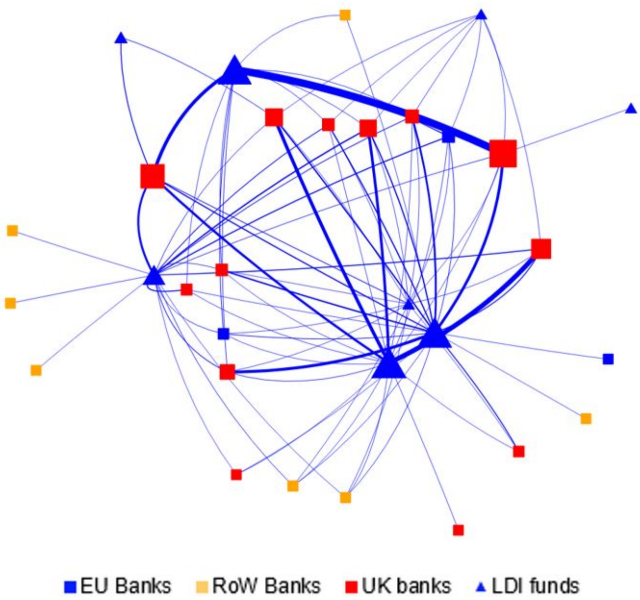


Figure 3 – Derivatives exposures of LDI funds (aggregated at manager level). The size of the node is proportional to gross derivatives exposures and the thickness of the edges is proportional to bilateral derivatives exposures. Source: EMIR and authors' calculation.

decline in NAV during the event. The mark-to-market value of IRDs exposed to floating rates, such as those used by LDI funds, plummeted as well, resulting in variation margins.

Overall, we estimate that liquidity demands from collateral requests and variation margins amounted to around €29bn, above available cash and MMFs held by LDI funds (Figure 4, right panel).[8] Therefore, LDI funds had to liquidate sovereign bonds to raise cash, thereby amplifying the downward price pressure on gilts (Pinter [2023])[9]. The price impact of LDI sales was also visible in the deterioration of market liquidity measured by bid-ask spreads (Figure 4, left panel) and the forced sales created a negative feedback loop: sales of gilts resulted in lower bond prices, triggering further liquidity demands and additional forced sales by LDI funds. The negative effect was also amplified by the fact that most gilts held by LDI funds were concentrated on the long-term and inflation-linked segment of the UK sovereign bond market (Dunne et al. [2023]). The forced deleveraging by LDI funds triggered the intervention of the Bank of England through temporary asset purchases targeted at the long end of the gilt market. Such intervention allowed yields to decline and provided time for LDI managers to reduce their risk profile.

Following the stress event, national competent authorities (NCAs) in the EU

8 We estimate variation margins by comparing the market value of LDI derivatives between 22 and 27 September as reported in EMIR. Collateral requests are estimated by applying observed change in yields to sovereign bonds used as collateral as of 22 September.

9 LDI funds also used recapitalisation (where pension funds are requested to inject capital into the funds) to raise cash, but operational challenges made the recapitalisation process complex and in some cases not possible (as recapitalisation typically occurs over one week), see Chen and Kemp [2023].

(a) 30Y GBP yields surged amid low liquidity (b) ...triggering large liquidity demands

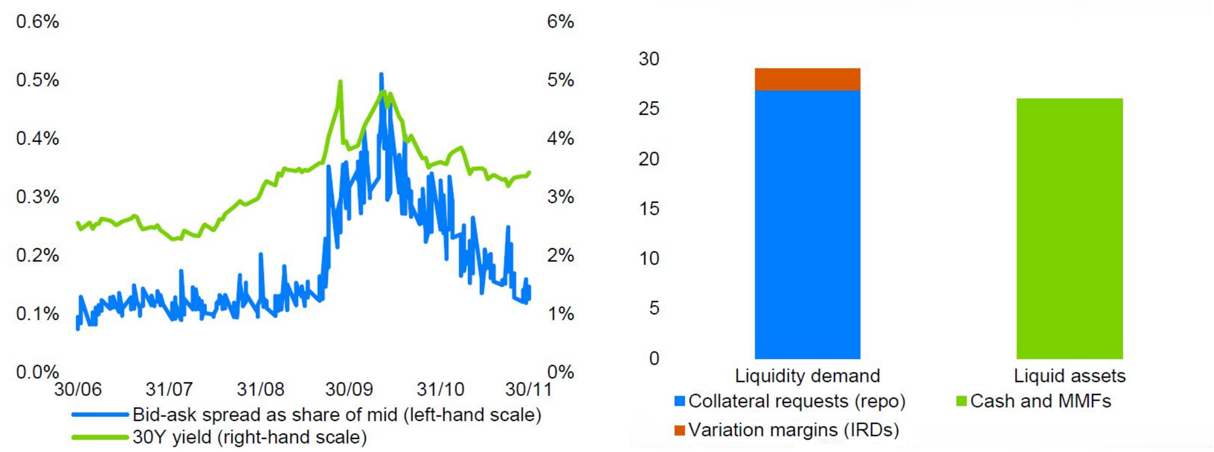


Figure 4 – UK 30 yields and liquidity demands for GBP LDI funds. Panel (a) displays bid ask spread on 30Y Gilt yields and levels. Source: Refinitiv Eikon, ESMA and ESRB. Panel (b) displays liquidity demand based on estimated changes in collateral value for repos and based on reported changes in market value for IRDs, in €billion. Changes between 22 and 27 September 2022 for collateral requests and 23 and 30 September for variation margins. Source: AIFMD, EMIR, SFTR and ESMA.

undertook a range of measures to ensure the resilience of GBP LDI funds. First, in November 2022, the Commission de Surveillance du Secteur Financier (CSSF) and the Central Bank of Ireland provided supervisory guidance to GBP LDI managers, by asking them to maintain the current level of resilience (measured by a fund's ability to withstand an interest rate shock of 300 to 400 basis points) and a reduced risk profile (ESMA [2022]). Second, in April 2024, both NCAs imposed investment restrictions under Article 25(3) of the AIFMD. Such restrictions require GBP LDI funds to be resilient to a shock of at least 300 basis points. The measure entered into force at end-July 2024 (ESMA [2024]).

5 Stress testing LDI funds in 2023

5.1 Overview

In the context of policy measures taken by NCAs, it is important to assess the resilience of GBP LDI funds. In particular, the 2022 LDI crisis has shown that when LDI funds are not able to meet liquidity demands stemming from interest rate shocks, they have to liquidate sovereign bonds, triggering a risk of a negative feedback loop. Therefore, we perform liquidity stress tests across our sample of funds.

Stress test scenarios

We use two stress test scenarios. In the first scenario, we assume a parallel shift in the GBP yield curve of 100 basis points, which is broadly commensurate with the observed shock during the mini-budget event. The second scenario features a 300 basis point shock, which relates to the expected level of resilience foreseen under the new investment restrictions for GBP LDI funds, and it is also close to the spike in long-term inflation yields during the September 2022 event.[10]

LDI exposures to IRDs and repo borrowing end-2023

Overall, LDI funds had derivatives exposures of around €130bn as of end-2023, consisting mostly of IRDs with the Sterling Overnight Index Swap (SONIA) as underlying (€65bn, around 61% of exposures), inflation swaps (€52bn, 35% of total exposures) and other derivatives. A large share of derivatives positions had a maturity above 10 years (46% of notional exposures and 55% in number of positions). Repo borrowing amounted to €93bn, almost exclusively backed by UK gilts.

Liquidity demands from repo and losses on bond portfolio

To assess the impact of the shock on bonds held and pledged by LDI funds, we need to estimate the change in the market value of those bonds for each fund. The impact of an interest rate shock Δr on the price of a bond P posted as collateral is given by

$$\frac{\Delta P}{P} \approx -D \cdot \Delta r + \frac{1}{2} \cdot C \cdot (\Delta r)^2$$

where D is the duration of the bond and C is the convexity.

We use SFTR data to identify the list of instruments pledged as collateral and we retrieve information on bond duration and convexity for each instrument using commercial data sources.

Liquidity demands related to variation margins on derivatives

For interest rate derivatives, estimating the impact of the shock is more complex. First, it requires mapping LDI exposures to interest rate derivatives using EMIR. Second, we need to price each derivative using a model that takes into account the characteristics of the contract (e.g. coupon frequency, maturity date, underlying rate). The pricing model can then be used to simulate the new price of the contract, after applying the interest rate shock to the discount rate and the forward curve (see Appendix A for further details). The difference between the two valuations provides an estimate of the mark-to-market impact of the shock on the value of the IRD and hence the variation margins faced by LDI funds (for a similar approach applied to Italian banks see Bianchi et al. [2024]).

Once liquidity demands are calculated from repo positions and IRDs, they are compared to available liquid assets. The liquid assets are comprised of cash and MMF shares, which can be mobilised immediately to post cash, and by unpledged sovereign bonds which can either be used as collateral for repo or sold to obtain

[10] Bandera and Stevens [2024] show that the increase in yields on 30-year inflation-linked bonds culminated at around 250 basis points compared to 130 basis points for 30-year nominal yields.

cash. In that context, the behaviour of LDI managers can transmit stress to the financial system. If MMF shares are redeemed, liquidity stress might spread to MMF managers who will in turn need to dispose of assets to meet investors' redemptions and this might contribute to liquidity issues in short-term funding markets. If sovereign bonds are sold, the absorption capacity of the market might be too limited to avoid further declines in bond prices.

5.2 Results

Scenario 1: 100 basis point shock

Under the first scenario, liquidity demands amount to around €12.5bn, stemming mostly from repo collateral requests (€10.9bn) and to a lesser extent from IRD margin calls (€1.9bn, Figure 5). Such liquidity demands are greater than the cash and MMF shares held by LDI funds, resulting in a liquidity shortfall of around €4.9bn. Funds would have to use unpledged bonds to meet the liquidity demands. In that case, the liquidity shortfall would reduce to 0.1bn, as most funds would be able to meet their liquidity demands. Under that scenario, assuming that LDI funds would need to dispose of gilt to raise cash, asset sales would amount to around €7.6bn. This volume of sales is large, as it amounts to around 55% of the average daily trading volumes of UK sovereign bonds.[11] LDI funds might avoid those forced sales, if unpledged bonds could be directly used as collateral for repo, or if other assets could be pledged.[12]

These stress results can be compared to the analysis carried out by ESRB [2023] which features a similar shock, using regulatory data before the stress event (June 2022). First, liquidity demands are lower end-2023 than in June 2022[13] (€13bn compared with €22bn). This partially reflects the deleveraging that took place between the two periods (Figure 1) with a reduction of the AuM to NAV ratio from 5.1 to 3.5. The decline relates to a reduction in repo borrowing from €141bn to €93bn and a reduction in OTC derivatives exposures from €178bn to €135bn. Relatedly, the liquidity shortfall would be lower in 2023 (€5bn compared with €7bn in June 2022) when considering only cash and MMF shares. However, the liquidity shortfall is higher in relative terms: whereas in June 2022 cash and MMFs could have been used to cover around 70% of liquidity demands (€15bn out of 22bn), at end-2023 those assets would only cover 60% of liquidity needs (€8bn out of 13bn). The amount of unpledged sovereign bonds that could be mobilised to meet liquidity demands is identical across both periods, at around €131bn.

---

11 According to the Bank of England letter to the Chair of the Treasury Committee, 5 October 2022, average daily trading volumes on UK sovereign bonds were around of GBP 12 billion (around €14 billion).

12 Tipping [2024] reports that pension schemes have explored using corporate bonds or fund shares as collateral for repo, although such practice could expose LDI funds to additional market risks on the collateral they use.

13 The ESRB analysis uses a slightly different sample of LDI funds that also include EU LDI funds. However, since most LDI funds are in GBP, the results are broadly comparable.

Scenario 2: 300 basis points shock

In the case of a very large shock, liquidity demands would amount to €31bn (including €26.9bn from collateral requests and 4.1bn in variation margins) and the liquidity shortfall would equal €17.1bn when considering cash and MMF shares, and €0.9bn when including the use of unpledged sovereign bonds (Figure 5). In the latter case, asset sales volumes would reach €14bn, more than the average daily trading volumes of UK sovereign bonds.

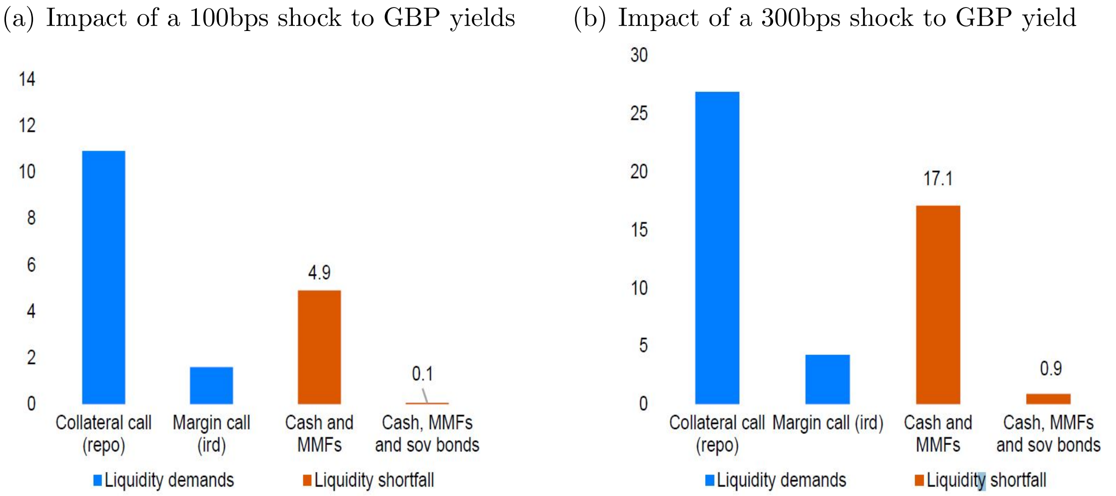


Figure 5 – Liquidity demand for 100 and 300 bps shock. Impact of a 100 (300) bps interest rate shock on pledged collateral for repo and liquiditiy shorfall in €billion. Sources: AIFMD, EMIR, SFTR and ESMA.

Counterparty analysis

We use the results of the second stress test scenario to map potential liquidity demands from LDI counterparties and measure the impact of a potential LDI default on banks. Figure 6 shows the liquidity demands that bank counterparties would request from LDI managers on behalf of their funds. For the three largest managers, liquidity demands would be spread out across multiple banks, with a concentration across a few banks. These banks would, in turn, be exposed to the largest LDI managers. The high degree of interconnectedness, largely reflecting the role played by a few large international banks in derivatives and repo markets, can be a further transmission channel between banks and non-banks.

6 Conclusion and extension of the work

We have shown how regulatory data at entity and activity level can be used to assess and monitor risks related to leverage for non banks through an application to EU GBP LDI funds. Our analysis could be extended across several dimensions. First, the framework could be used to perform reverse stress tests, to identify the size of the interest rate shock that would exhaust available liquidity for each fund, similarly to ESRB [2023]. Second, the analysis could be applied to other types of

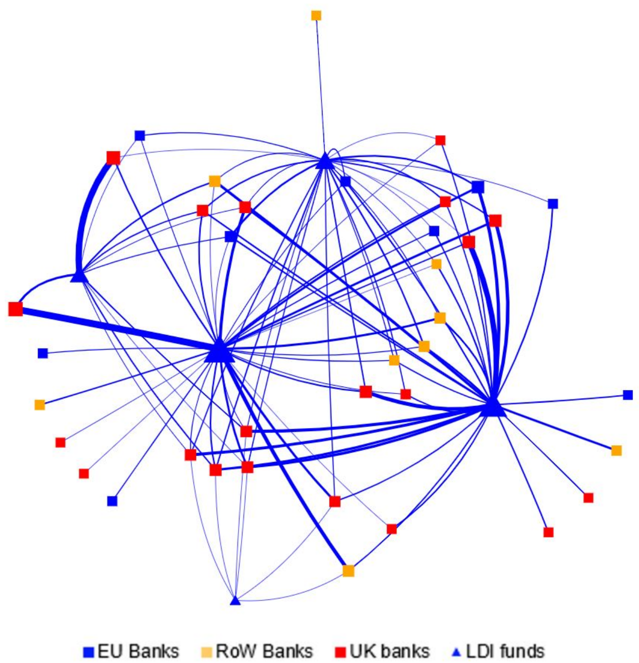


Figure 6 – Collateral demand following the 300bps shock. The size of the node is proportional to total collateral demand and the thickness of the edges is proportional to bilateral collateral demand. Source: EMIR and authors' calculations.

entities, provided that balance sheet information and exposures data are available. Finally, the tools developed here could be mobilised to perform sector-wide or system-wide stress testing.

Bibliography

Bandera, N. and J. Stevens (2024): "Monetary policy consequences of financial stability interventions: assessing the UK LDI crisis and the central bank policy response," *Bank of England, Staff Working Paper*.

Barria, R. and G. Pinter (2023): "Mispricing in inflation markets," *Bank of England, Staff Working Paper*, https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2023/mispricing-in-inflation-markets.pdf.

Bianchi, M., D. Ruzzi, and A. Segura (2024): "Shifting the yield curve for fixed-income and derivatives portfolio," https://arxiv.org/abs/2412.15986.

Bouveret, A. and L. Danieli (2021): "Vulnerabilities in money market funds," in *ESMA Report on Trends, Risks and Vulnerabilities*, No. 1, https://www.esma.europa.eu/sites/default/files/trv_2021_1-vulnerabilities_in_money_market_funds.pdf.

Bouveret, A. and M. Haferkorn (2023): "Leverage and derivatives: The case of Archegos," *Journal of Securities Operations & Custody*, 15, 238-250.

Chen, R. and E. Kemp (2023): "Putting out the NBFire: Lessons from the UK's liability-driven investment (LDI) crisis," in *IMF Working Paper*, No 23/210, https://www.imf.org/en/Publications/WP/Issues/2023/09/29/Putting-Out-the-NBFIRE-Lessons-from-the-UK-s-Liability-Driven-Investment-LDI-Crisis-539683.

Dunne, P., A. Ghiselli, F. Ledoux, and B. McCarthy (2023): "Irish-Resident LDI funds and the 2022 Gilt Market Crisis," in *Central Bank of Ireland, Financial Stability Notes*, No 7, https://www.centralbank.ie/docs/default-source/publications/financial-stability-notes/irish-resident-ldi-funds-and-the-2022-gilt-market-crisis.pdf?sfvrsn=f26c9c1d_8.

ESMA (2022): "ESMA welcomes NCAs' work to maintain resilience of liability driven investment funds," https://www.esma.europa.eu/press-news/esma-news/esma-welcomes-ncas%E2%80%99-work-maintain-resilience-liability-driven-investment-funds.

——— (2024): "Advice of the European Securities and Markets Authority of 26 April 2024 on a proposed measure by the Commission de Surveillance du Secteur Financier under Article 25 of Directive 2011/61/EU," https://www.esma.europ a.eu/sites/default/files/2024-04/ESMA50-43599798-9492_Advice_under_article_25_AIFMD_CSSF_measure.pdf.

ESRB (2023): "EU Non-bank Financial Intermediation Risk Monitor 2023," https://www.esrb.europa.eu/pub/pdf/reports/nbfi_monitor/esrb.nbfi202306~58b19c8627.en.pdf.

Girardi, G., K. Hanley, S. Nikolova, L. Pelizzon, and M. Getmansky Sherman (2021): "Portfolio similarity and asset liquidation in the insurance industry," *Journal of Financial Economics*, 142, 69-96.

Pinter, G. (2023): "An anatomy of the 2022 gilt market crisis," *Bank of England, Staff Working Paper*, https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2023/an-anatomy-of-the-2022-gilt-market-crisis.pdf.

Tipping, N. (2024): "Nervous UK pension schemes want liquidity fixes – it'll cost them," *Risk.net*, https://www.risk.net/investing/7959353/nervous-uk-pension-schemes-want-liquidity-fixes-itll-cost-them.

Appendix

A Pricing interest rate derivatives using EMIR data

A typical interest rate swap is a financial contract where one counterparty agrees to pay a floating rate to another counterparty who commits to pay a fixed rate in return over the maturity of the contract.

Below, we take the point of view of the buyer who wants to secure fixed cash flows over a long time horizon T, like LDI funds. The buyer receives fixed cash flows and pays a floating rate in return. For the fixed leg, the future cash flows are defined as the sum of the present value of the fixed payments, equal to the agreed fixed rate r multiplied by the notional of the contract N, discounted by a risk-free rate (the discount factor, noted δ). For the floating leg, the cash flows are equal to the future floating rate f (derived from the forward curve) multiplied by the notional and the discount factor. The value of the swap, from a buyer point's of view, is given by the value of the fixed leg (cash inflows) minus the value of the floating leg (cash outflows).

$$\text{IR Swap} = \underbrace{\sum_{t=0}^{T} r\, N\, \delta_t}_{\text{fixed leg}} - \underbrace{\sum_{t=0}^{T} f_t\, N\, \delta_t}_{\text{floating leg}}$$

Discount factor and forward curve

We use Overnight Index Swaps (OIS) as risk-free rates for the discount factor. For pound-denominated assets we take the swaps on the Sterling Overnight Index Average (SONIA) as the risk-free rate s_t and we calculate the discount factor based on a continuously compounded rate.

$$\delta_t = e^{-t\, s_t}$$

The forward rate f_t between two periods t_1 and t_2 is defined as follows:

$$f_t = \left(\frac{\delta_{t_1}}{\delta_{t_2}}\right)^{\frac{1}{(t_2 - t_1)}} - 1$$

The forward curve of the floating rate is time-dependent and changes with market conditions. It represents the so-called "market expectations".

The discount factor curve depends on the maturity dates of the instrument used as benchmark for the risk-free rate. However, the future cash flows of the fixed and floating legs are not necessarily scheduled on the same dates. Therefore, the relevant discount factor curve needs to be interpolated to obtain a vector of the discount factors for each cash flow date. We use cubic splines to interpolate the curve between each observation.

In accordance with industry practices the results of the subtraction of two dates depend on the day count convention agreed in the derivative contract and is expressed in a fraction of a year.

Stress test

The shock on the floating leg is two-fold, first on the forward floating rate curve and second on the discount factor. A shock on the discount factor reflects a change in money market conditions and interest rate expectations, for instance following a central bank decision on key interest rates. The shock on the forward floating rate curve can have the same origin, especially if the floating rate is also the benchmark used for the discount factor. Shocks to the floating rate can also reflect changes in market expectations of other variables such as inflation expectations, in particular when the floating rate is related to inflation rates, as is the case for inflation swaps. In contrast, the shock to the fixed leg is only driven by the shock to the discount factor.

Data cleaning

Data quality issues might arise in EMIR, including imperfect reporting of the floating rate of the swap and missing data. Hence, we have made some assumptions in order to harmonise the data and increase its usability. We have implemented the following rules:

- Ambiguous benchmark rate for floating rate payment. The following strings are replaced by “SONIA”: GBP-SONIA-COMPOUND, SONIO/N Index. For other floating rate references, we use an enhanced Ratcliff-Obershelp algorithm to compare them to SONIA, relying on the difflib library in python. If they are deemed close enough they are replace by “SONIA”.
- Missing reference period of the floating rate. Since the reference period for SONIA is a day, other misreported values are replaced by day.
- Missing day count convention. If the day count convention is not reported, by default we assume it is Actual/365.

Implementation

A dedicated python library has been developed to price interest swap derivative contracts. It implements the methodology and rules described above. The library gives a valuation to the floating and the fixed legs separately, and then use the two values to obtain the value of the contract depending on whether the report has been sent by the buyer or by the seller. The library relies on discount and OIS curves retrieved from Eikon (Reuters). The library is used to shock the contracts by shifting the discount curves and the floating rate upwards or downwards. Additionally, the library produces plots to illustrate the impact of the shocks, see Figure A.1 and Figure A.2.

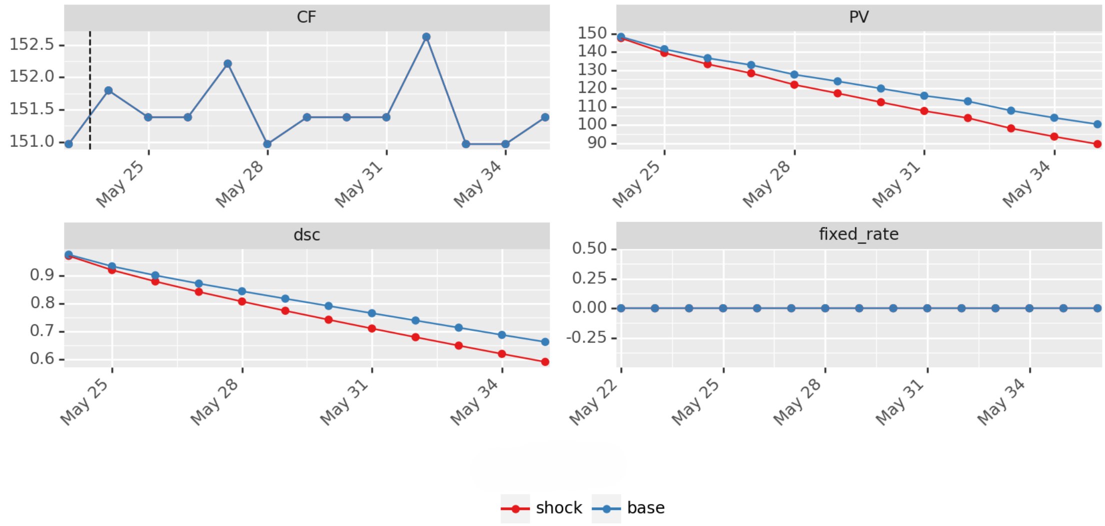


Figure A.1 – Valuation of a fixed leg following a shock of 100bp. Valuation of the fixed leg (receiving twice a year a fixed rate equal to 1.74 bps) of a contract starting in January 18^{th} 2022 with maturity May 29^{th} 2035 with notional equal to 8.7 mln.

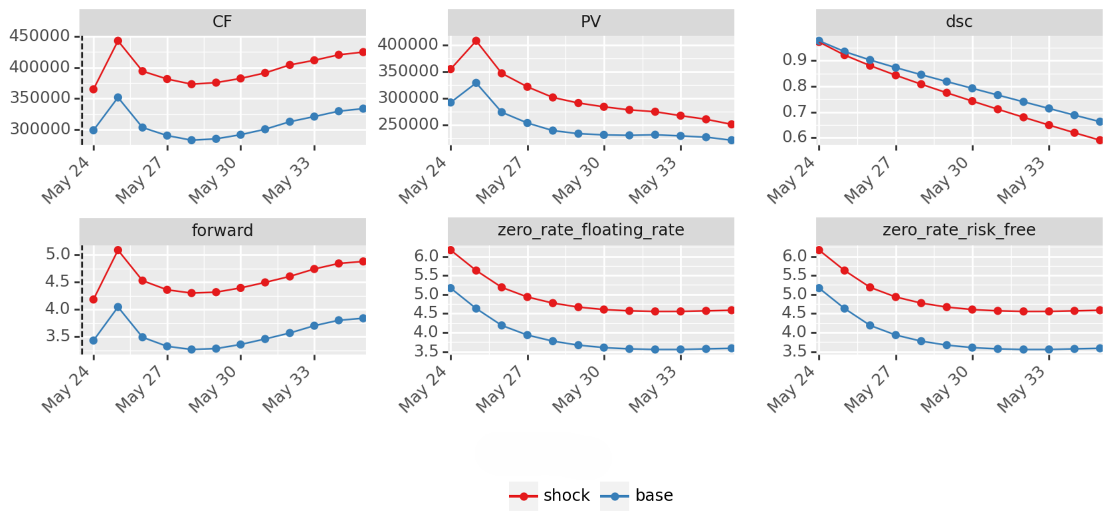


Figure A.2 – Valuation of a floating leg following a shock of 100bps. Valuation of the floating leg (receiving twice a year the SONIA rate) of a contract starting in January 18^{th} 2022 with maturity May 29^{th} 2035 with notional equal to 8.7 mln.

New NBFI monitoring tool

Annalaura Ianiro,[a,1] Linda Rousová,[a] and Dario Ruzzi[b]

[a]*Directorate General Macroprudential Policy and Financial Stability, European Central Bank, Frankfurt am Main, Germany*
[b]*Financial Stability Directorate, Banca d'Italia, Rome, Italy*

Abstract. Recent episodes of market stress, such as the Archegos collapse in 2021 and the UK Gilt market turmoil, have highlighted the systemic vulnerabilities posed by leveraged non-bank financial institutions. These events showed how non-bank entities can accumulate opaque, leveraged exposures through derivatives, amplifying risks and market volatility. To address these challenges, we introduce an advanced monitoring tool designed to detect and evaluate risks associated with leveraged non-banks. Initially focused on equity derivative markets, this modular tool utilizes transaction-level data from the European Market Infrastructure Regulation (EMIR), complemented by other sources, to assess key risk dimensions, including size, concentration, synthetic leverage, and liquidity preparedness. By filling critical gaps in risk assessment and transparency, this tool aims to support proactive risk mitigation, aligning with global efforts to enhance financial stability in increasingly complex markets.

Key words: non-bank financial institutions, systemic risk, leverage, liquidity, data gaps.

[1] This publication should not be reported as representing the views of the European Central Bank and Banca d'Italia. The views expressed are those of the authors and do not necessarily reflect those of the European Central Bank and Banca d'Italia. The authors are grateful to the participants at the 2024 EMIR Workshop of the Banca d'Italia and of internal ECB's seminars for their helpful comments.

1 Introduction

The non-bank financial intermediation (NBFI) sector has become pivotal in global financial markets, although some of its entities – such as family offices – operate outside of the traditional regulatory perimeter [FSC, 2024]. While the operational flexibility enables innovation and efficiency, the sector also introduces unique vulnerabilities, particularly when non-banks engage in complex, leveraged financial practices. A prominent example of this is non-banks' use of derivatives to build synthetic leverage, which can amplify both returns and risks.

Recent market events, such as the Archegos collapse in 2021 and the turmoil in the UK Gilt market in 2022, have highlighted the systemic risks associated with leveraged non-banks. Archegos, a U.S.-based family office, defaulted on margin calls on its equity total return swap (TRS) portfolio, triggering multi-billion-dollar losses for its counterparties. In particular, Credit Suisse suffered a $5.5 billion loss, the largest from a single default event in its 165-year history, which contributed to its demise in 2023 (see ESMA [2022]). Similarly, the stress in the UK Gilt market revealed how a sudden liquidity shock can destabilize non-bank players with leveraged exposures and how such stress can spillover to the broader financial system, requiring public interventions (see Chen and Kemp [2023]). While these episodes did not trigger broader financial crises, they highlighted the potential for the NBFI sector to amplify market volatility and disrupt financial stability.

Despite the regulatory reforms enacted in the wake of the global financial crisis, significant gaps remain in the oversight and transparency of non-bank entities. These gaps are particularly pronounced for those employing derivatives to build up leverage. Current frameworks often fail to provide regulators with timely and granular insights, making it difficult to detect emerging risks before they escalate.

This paper introduces a novel monitoring tool designed to help address these challenges. Initially tailored for leveraged non-banks active in equity derivative markets, the tool employs a modular structure, allowing for future extensions to other asset classes. By using detailed transaction-level data from the European Market Infrastructure Regulation (EMIR) and other granular data sources, the tool enables comprehensive assessment of risk dimensions such as size, concentration, synthetic leverage, and liquidity preparedness. Through its robust design, the tool enhances regulators' ability to monitor systemic risks and take proactive measures to mitigate them.

This initiative aligns with broader global efforts to strengthen the resilience of the NBFI sector, building on the work of institutions such as the Financial Stability Board (FSB) and the International Organization of Securities Commissions (IOSCO).[2] By addressing critical vulnerabilities, the proposed tool seeks to fill a key gap in the regulatory landscape, supporting financial stability in an increasingly complex market environment.

[2] See for example Cheng et al. [2023].

2 Data

2.1 Source

The NBFI monitoring tool is built using data collected under EMIR. This regulation was introduced in 2012 in response to the global financial crisis, with the aim to enhance transparency, mitigate systemic risks, and prevent market abuse. The regulation mandates that entities in the EU engaging in derivative transactions report them to trade repositories (TRs). The data thus include granular, transaction-level information such as counterparty details, contract specifics, product classifications, underlying assets, notional amounts, contract valuations, margins, and various asset-class-specific variables. Unlike the US Dodd-Frank Act, EMIR also covers exchange-traded derivatives and entails double-sided reporting, meaning both counterparties to a trade, if subject to reporting, must report.

These data are critical for a whole range of authorities, including the European Central Bank (ECB), to monitor financial stability and assess systemic risks. However, the usability of EMIR data has been significantly hampered by persistent challenges in reporting and data quality. EMIR Refit, which entered into force on 29 April 2024, helps overcome some of these issues, but there is still ample room for improvement. Therefore, substantial efforts in data cleaning, standardization, and IT infrastructure development remain necessary to fully reap the benefits from this critical data source.

2.2 EMIR cleaning procedure

Due to their volume, velocity and variety, EMIR data can be classified as "big data", posing significant challenges for using them. To address these complexities, between 2017-2018 the ECB, in collaboration with the European Systemic Risk Board (ESRB) Secretariat, built a dedicated IT infrastructure to collect and process EMIR data. The EMIR IT system collects the daily information from the European Securities and Markets Authority (ESMA) TRACE platform[3] and applies a series of data processing steps. These include format conversion, enrichment with reference databases, data quality assurance, and de-duplication of double-sided trades. The data are then made available to internal ECB and ESRB Secretariat users via the ECB cloud-based analytical platform (see Agostoni et al. [2024] and Boneva et al. [2019]).

The ECB's Directorate General Macroprudential Policy and Financial Stability (DG-MF) developed an extensive cleaning pipeline (see Figure 1) which, starting from the data produced by the EMIR IT system, produces EMIR tables suitable for risk monitoring and analysis for financial stability purposes. Prior to the implementation of EMIR Refit, the cleaning pipeline was fully automated, producing daily updates stored on the ECB's cloud-based big data platform. However, to accommodate the structural changes introduced by EMIR Refit, the process is undergoing enhancements to handle the increased data complexity effectively. DG-MF's cleaning procedure consists of three steps:

[3] The ESMA IT infrastructure through which EMIR data are collected by TRs and made available to authorities.

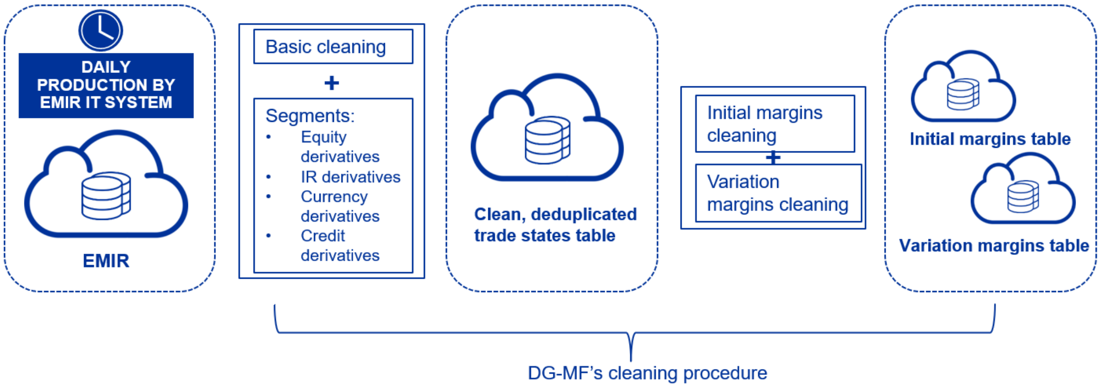


Figure 1 – DG-MF's EMIR cleaning procedure.

Step 1: Basic Cleaning

This phase addresses general issues across all asset classes. Key elements include:

- classification of contract type and asset class using the Classification of Financial Instrument (CFI) code;
- correction of execution timestamps and maturity dates;
- identification of implausible values for notional amounts through threshold settings;
- implementation of data quality flags to detect inconsistencies.

Step 2: Asset Class-Specific Segmentation

This phase employs tailored modules for different asset classes:

- interest rate derivatives: creation of tenor curves, validation of date consistency, and construction of a detailed product taxonomy;
- currency derivatives: development of a product taxonomy for currency forwards;
- equity derivatives: mapping of various ISINs or proprietary codes to the underlying ISIN and classification of options into European or American types;
- credit default swaps (CDS): further cleaning of notional values and identification of single-name, index, and multi-name CDS using external mapping data.

Step 3: Margins[4]

This final step reconciles collateralisation and margining data. Key activities include:

[4] This module was developed in collaboration with the ECB's Directorate General Market Infrastructure and Payments.

- identification and resolution of inconsistencies in reported collateral fields;
- assignment of unique portfolio codes for collateral reconciliation;
- creation of portfolio-level initial and variation margin tables, offering a clearer view of collateral requirements.

The process delivers a clean, enriched dataset that serves as a robust foundation for the new NBFI monitoring tool, delivering both precision and scalability to support current and future use cases.

Sector Enrichment: To complement the cleaning process, the dataset undergoes a sector enrichment step. Developed by Lenoci and Letizia [2021], this process employs hierarchical and incremental classification techniques, drawing on multiple external sources to enhance the granularity and accuracy of self-reported sector classifications within the EMIR dataset.

3 Tool

The monitoring tool integrates seamlessly with the cleaned and enriched EMIR data produced for financial stability purposes. Flexible scripts are employed to extract the required information based on parameters such as asset class and contract type. These scripts generate tables populated with risk indicators tailored to individual entities, offering a granular view of exposures and associated risks (see Figure 2). Furthermore, its modular and scalable infrastructure ensures that the tool can incorporate additional data sources, such as Refinitiv Lipper[5], on individual investment funds, (already included) or Security Financing Transaction Regulation (SFTR) data, providing a framework for accommodating future regulatory and analytical requirements.

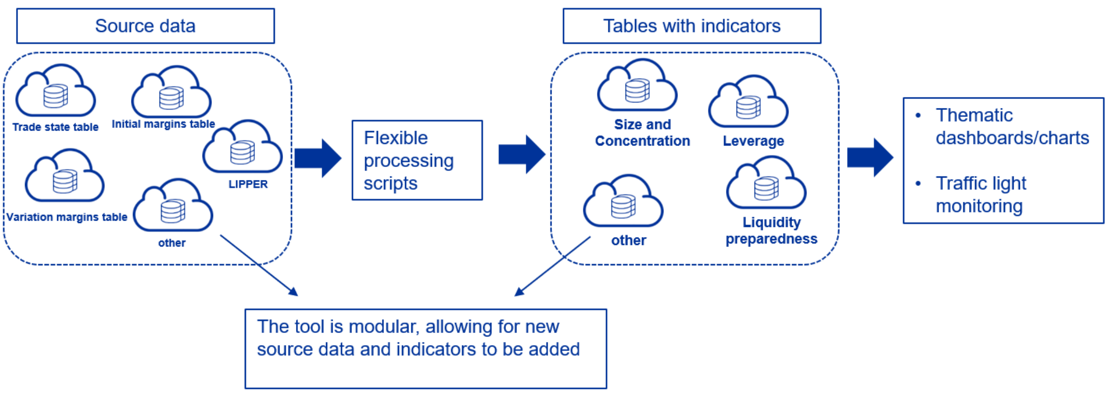


Figure 2 – NBFI monitoring tool's infrastructure.

5 Refinitiv Lipper is a commercial data provider offering granular investment fund data, including static information (e.g., domicile, asset class, investment strategy) and time series (e.g., full portfolio holdings, performance measures, total assets, net estimated flows, derived aggregated asset allocations).

3.1 Scope

The monitoring tool currently focuses on equity derivatives, encompassing all contract types. It includes non-banks with total equity derivatives exposures exceeding €1 billion in gross outstanding notional value. The tool provides monthly snapshots of relevant indicators, capturing market developments and allowing for timely identification of potential risks.

3.2 Indicators

The tool's core functionality lies in its ability to provide risk indicators that directly align with key risk dimensions. These indicators not only offer insights into current market conditions but also facilitate trend analysis by including one-month growth rates. The available indicators are shown in Table 1.

Size and Concentration	**Synthetic Leverage**	**Liquidity Preparedness**[a]
• Notional value (gross, long, and short) • Market share • Broker concentration (HHI[b], NHHI[c], cumulative top 3 brokers' share) • Underlying concentration (HHI, NHHI, cumulative top 3 underlyings' share)	• Gross notional value / IM • Gross notional value / total net assets	• Stock IM / cash • Flow (IM + VM) / cash • Stock IM / HQLA • Flow (IM + VM) / HQLA • Historical peak VM calls / cash buffer

[a] Entity-level data on liquidity buffers (from Refinitiv Lipper) is only available for investment funds.
[b] Herfindahl–Hirschman index.
[c] Normalised Herfindahl–Hirschman index.

Table 1 – NBFI monitoring tool's risk indicators.

Each group of indicators speaks to a different risk dimension. The size and concentration indicators give information on the overall scale of a non-bank's derivative exposure, as well as its market share and concentration in terms of broker counterparties or underlying assets. The synthetic leverage indicators assess the extent of leverage employed relative to the collateral provided in the form of initial margins (IM) and in the context of an entity's financial capacity. Finally, the liquidity preparedness indicators help to gauge an entity's ability to manage both expected and unexpected cash demands, crucial during episodes of market stress. As said, the latter indicators can be particularly useful in periods of market stress and, importantly, the tool's data collection and processing frequency can be shifted from monthly to daily to allow for more timely monitoring.

3.3 Summary output

The monitoring tool delivers insights on NBFIs' potential vulnerabilities through two key outputs: thematic dashboards and a traffic light framework. These outputs, currently under development, will allow users to visualize and interpret risk indicators effectively, facilitating early identification of vulnerabilities and informed decision-making.

Thematic Dashboards

The dashboards serve as the primary interface to explore and interpret the tool's risk measures. They provide dynamic, user-friendly visualizations, allowing for a detailed examination of individual indicators and their evolution over time. For example:

- charts combining exposure size with growth rates enable the identification of entities with significant exposures that are rapidly increasing (see Figure 3a);
- concentration-focused visualizations highlight whether exposures are clustered across brokers or underlying assets (see Figure 3b).

These visualisations can help present complex and big data in an intuitive format, aiding in the swift identification of emerging trends and potential criticalities.

(a) GNV (x-axis) and GNV 1-month (y-axis), € billions

(b) broker NHHI (x-axis) and underlying NHHI (y-axis)

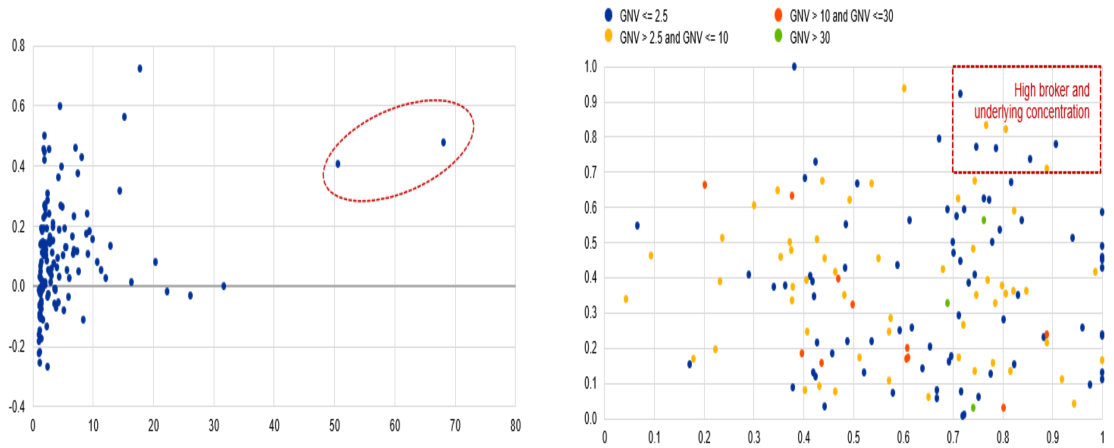


Figure 3 – EMIR data and authors' calculations. Q4 2023. Example charts of the thematic dashboards. Each point is the result of the aggregation of three entities. GNV stands for gross notional value, NHHI stands for Normalised Herfindahl–Hirschman index.

Traffic Light Framework

The traffic light framework aims to offer a clear, visually intuitive representation of risk levels across monitored entities. Figure 4 shows a mock-up. The idea is to assign different scores (colours) to entities, corresponding to a different percentile in the distribution of each risk indicator, as follows:

- **green:** the risk indicator is below the 25th percentile;
- **yellow:** the risk indicator is between the 25th and 75th percentile;
- **orange:** the risk indicator is between the 75th and 90th percentile;
- **red:** the risk indicator is above the 90th percentile.

The scores are then weighted together to obtain an overall risk score. The aggregation methodology is still under development, as more complex techniques than equal-weight average are being considered.

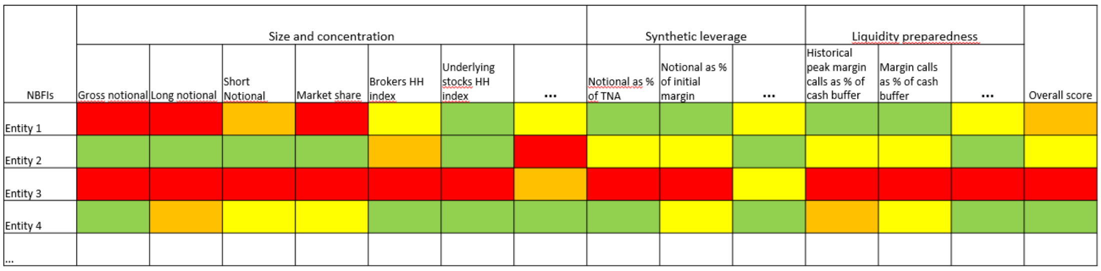


Figure 4 – Mock-up of the traffic light framework.

4 Concluding remarks

The increasing footprint of non-bank financial institutions in global markets highlights both their potential to foster financial innovation and their capacity to amplify systemic risks. The episodes of stress involving leveraged non-banks, such as the Archegos collapse and the UK Gilt market turmoil, underscore the urgent need for more effective monitoring and risk assessment tools.

This paper presents a novel NBFI monitoring tool designed to bridge gaps in risk assessment and data transparency. By leveraging on granular transaction level data from EMIR and integrating additional data sources, the tool provides a comprehensive and scalable framework to analyze key risk dimensions, including size, concentration, leverage, and liquidity preparedness. The modular structure of the tool allows for future expansion to other asset classes and regulatory domains, ensuring its relevance in an evolving financial landscape.

The insights generated by this tool not only have the potential to enhance the ability of central banks to detect emerging vulnerabilities but also to support the development of targeted interventions to mitigate systemic risks. The traffic light framework and thematic dashboards, currently under development, aim to deliver useful information in an intuitive and user-friendly format, facilitating timely and informed decision-making.

However, the tool's effectiveness is inherently tied to the quality and scope of the underlying data. While EMIR data offer significant insights into derivative markets, there remain critical gaps in the granularity and coverage of data on non-bank financial entities. A full understanding of the risk exposures and interconnections of these entities requires broader and higher-quality data. Addressing these limitations will necessitate concerted efforts to improve data collection frameworks, harmonize reporting standards, and enhance collaboration, as well as data sharing, across jurisdictions and regulatory bodies (see FSC [2024]).

Looking ahead, the tool's functionalities must continue to evolve to meet the growing complexity of financial markets. Enhancing its analytical capabilities, integrating additional data sources, and refining its outputs will be critical to ensuring its relevance and effectiveness. Simultaneously, addressing the limitations in data quality and granularity should remain a priority. By building on these improvements, the tool has the potential to enhance significantly the understanding of leveraged NBFIs, supporting proactive risk assessment in an increasingly dynamic financial landscape.

Bibliography

Agostoni, G., A. Ianiro, A. Jukonis, F. Lenoci, E. Letizia, and G. Skrzypczynski (2024): "Using trade-level derivatives data for macroprudential analysis," in *64th ISI World Statistics Congress*, https://isi-web.org/sites/default/files/2024-03/ottawa-2023_ips-1077-Grzegorz%20Skrzypczynski.pdf.

Boneva, L., B. Böninghausen, L. Rousová, and E. Letizia (2019): "Derivatives transactions data and their use in central bank analysis," in *ECB Economic Bulletin*, 6, https://www.ecb.europa.eu/press/economic-bulletin/articles/2019/html/ecb.ebart201906_01~dd0cd7f942.en.html.

Chen, R. and E. Kemp (2023): "Putting out the NBFire: Lessons from the UK's liability-driven investment (LDI) crisis," in *IMF Working Paper*, vol. No 23/210, https://www.imf.org/en/Publications/WP/Issues/2023/09/29/Putting-Out-the-NBFIRE-Lessons-from-the-UK-s-Liability-Driven-Investment-LDI-Crisis-539683.

Cheng, K., Z. Liu, S. Pezzini, and L. Yu (2023): "Building an integrated surveillance framework for highly leveraged NBFIs - lessons from the HKMA," https://www.bis.org/publ/bispap137.htm.

ESMA (2022): "Leverage and Derivatives - the Case of Archegos," https://www.esma.europa.eu/sites/default/files/library/esma50-165-2096_leverage_and_derivatives_the_case_of_archegos.pdf.

FSC (2024): "Eurosystem response to EU Commission's consultation on macroprudential policies for non-bank financial intermediation (NBFI)," https://www.ecb.europa.eu/pub/pdf/other/ecb.eurosystem_response_EUcommission_on_macroprudential_policies_NBFI_202411~a38ef4423d.en.pdf.

Lenoci, F. and E. Letizia (2021): "Classifying counterparty sector in EMIR data," in *Data Science for Economics and Finance*, Springer, 117-143.

The effects of obligatory central clearing for Dutch pension funds

Marnix Driesse[a,1] and Ellen van der Woerd[a]

[a] *Payments, Cash & Market Infrastructure, De Nederlandsche Bank, Amsterdam, the Netherlands*

Abstract. We analyze the effect of the central clearing obligation, which came into effect in June 2023, on Dutch pension funds using the EMIR data. This dataset gives unique insight into the clearing behavior of pension funds as standard supervisory data do not contain information on clearing. We deduplicate the data on counterparty level and use quarterly supervisory data as a quality check. We find that pension funds mainly use Euro denominated interest rate swaps to hedge their interest rate risk. The clearing rate for these types of derivatives has steadily increased since 2021 in preparation for mandatory clearing, with an increase from 18% to 45% for the larger pension funds between 2019 and 2024. Moreover, we find that more counterparties became available as possible clearing members for pension funds, leading to a less concentrated market.

Key words: Pension funds, derivatives, clearing obligation, EMIR.

1 The authors are grateful to Vera Boots, Oana Furtuna, Ilaria Gianstefani and Tom van der Zanden for their helpful suggestions. This publication should not be reported as representing the views of De Nederlandsche Bank. The views expressed are those of the authors and do not necessarily reflect those of De Nederlandsche Bank.

1 Introduction

With the introduction of the European Market Infrastructure Regulation (EMIR) in 2012 (European Union [2012]) came the obligation to centrally clear for counterparties with large derivative portfolios. For a long time, pension funds were exempt from this obligation for two reasons. First, pension funds have by nature a large one-sided position, as they use the derivative market mainly to hedge their interest rate risk. These large positions may not be very appealing to clearing members, as the resulting expansion of their balance sheet leads to higher capital requirements and credit risk in the case of default. Therefore, there was concern whether clearing members would be willing to accept pension funds as clients. Secondly, the large one-sided positions can at times lead to large margin calls. Pension funds might not be able to pay these margin calls in a timely manner, due to the lack of liquid resources.

After a decade, the exemption was lifted on June 18, 2023, as both pension funds and the market had enough time to adapt to the new situation and find strategies to deal with the above mentioned issues. In this analysis we investigate the changes in the derivative market for Dutch pension funds, with a focus on euro denominated interest rate swaps (EUR IRS), as the Dutch market has around thirty-five pension funds with derivative portfolio's large enough to fall under the clearing obligation.[2] Using the EMIR data, we analyze the change in central clearing rate and clearing member availability due to the mandatory clearing requirement. This dataset provides unique insight into the clearing behavior of pension funds as standard supervisory data on pension funds do not contain information on whether certain derivative contracts are cleared. Moreover, the EMIR data offers detailed information on derivative positions at the contract level, enabling the extraction of insights regarding counterparties, rates, and margins.

The results of our analysis are interesting for policy makers, financial stability experts as well as pension fund supervisors. For policymakers, this analysis provides valuable insights into the effects of mandatory clearing and can provide guidance for future policy decisions. Due to the size of the Dutch pension fund sector, the impact on financial stability might be relevant if there were fire sales of assets. If several pension funds had to sell assets simultaneously to meet margin calls, it could create pressure on the financial markets. For supervisors, the result of this analysis is relevant for the supervision of the mandatory clearing obligation and the general derivative exposures of pension funds.

In this paper we first discuss our data cleaning process in Section 2, with a focus on the method for deduplication of the data. We show how comparison with an independent data source, such as supervisory data, leads to an informed data cleaning procedure. This is followed by our results in Section 3. In the closing section we conclude and outline the next steps.

[2] Pension funds are subject to the clearing obligation when their total position in interest rate derivatives exceeds 3 billion Euro (ESMA [2019]).

2 Data quality and supervisory data comparison

Data quality is a common topic of discussion when working with EMIR data. Choices being made in the data cleaning and deduplucation process potentially can impact the results significantly. In this section we show the importance of having an independent data source to make informed decisions about the data cleaning process. We start this section by providing details on our general data cleaning process and then showcase how we determine our deduplication procedure based on a comparison with supervisory data.

Our dataset contains EMIR data from Q1 2019 until Q2 2024 (until EMIR REFIT at the end of April 2024). We find that data prior to this date tend to be less reliable, while at the same time we want to include data from well before the clearing obligation to fully capture the effects of the obligation. We only include dates on which all trade repositories have submitted a report.[3] To select data related to pension funds we filter for trades where either the reporting or other counterparty has a legal entity identifier (LEI) code that is included in the list of supervised Dutch pension funds (DNB [2023]). Furthermore, we only include pension funds that were active in the derivative market on 29 December 2023.

Within EMIR data, it is possible to assess data quality by comparing contracts reported by both counterparties, which is the case when both counterparties reside in the European Union (EU). In contrast, contracts involving a counterparty outside the EU are only reported by the pension fund. In the case of perfect data quality, all trade IDs for contracts between EU parties should be reported twice, and the notional amounts reported by pension funds and their counterparties should match. We find that for most time periods around 80% of the trade IDs match. Furthermore, we find that when the trade ID matches, the total notional matches as well, indicating good data quality of the notional field.

Although 80% is a relatively high matching percentage,[4] it indicates that it is not possible to use the trade ID for deduplication, as this would lead to double counting for 20% of outstanding contracts. Therefore, we deduplicate the data on counterparty level. We distinguish two different deduplicated sets. The first set contains all contracts reported by pension funds themselves. The other set contains all contracts reported by the counterparty, supplemented with contracts reported by the pension fund when the counterparty resides outside the EU and does not report.

To make an informed decision on how to deduplicate the data, we compare the EMIR data from 29 December 2023, with quarterly supervisory data (FTK[5]). DNB collects FTK data to support the regulatory supervision of Dutch pension funds. These data provide essential information to monitor financial health, assess solvency, and ensure adequate funding levels of the pension funds. DNB also uses the reports to evaluate the funds' resilience to market fluctuations, interest rate changes, and other economic factors. Dutch pension funds submit these data

[3] DNB sources EMIR data from CSV files instead of XML files, leading to some additional data quality challenges, including the identification of the correct business date.

[4] See Perez-Duarte and Skrzypczynski [2019].

[5] The Financieel Toetsingskader (Financial Assessment Framework), part of the Dutch Pension Act, establishes the legal financial requirements for pension funds.

quarterly, aggregated at the total level for each asset class, rather than at the individual contract level as done in EMIR. Characteristics of the data sources are shown in Table 1.

Data source	Granularity	Frequency	Reported by
EMIR - pension fund	Contract level	daily	Pension fund
EMIR - counterparty	Contract level	daily	Counterparty and pension fund in case counterparty does not report
Supervision data	Totals per pension fund per asset class	quarterly	Pension fund

Table 1 – Data sources.

To get a broad overview of the data quality we compare the total notional reported in the supervisory data for all pension funds with the totals reported in EMIR for both deduplicated sets. Figure 1 shows the results for both the total derivative portfolio and IRS. We find a significant over-reporting in EMIR data, especially for all asset classes combined. The discrepancy between the two EMIR datasets is primarily due to over-reporting by pension funds in the bilateral segment, rather than differing notional values reported for the same contract.

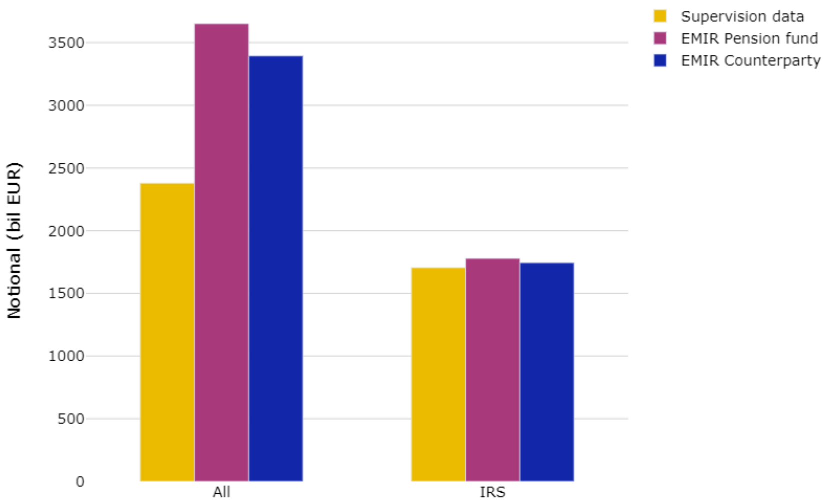


Figure 1 – Total outstanding notional on 29-12-2023 for the total portfolio and IRS. Sources are FTK data and EMIR data, as reported both by the pension funds and their counterparties.

To explain the over-reporting by pension funds, we look at when the contract was last evaluated. The left panel of Figure 2 shows that EMIR data contain a large number of contracts with a valuation date of more than 2000 days in the past. These are mainly contracts with non-EU counterparties, causing over-reporting both for the pension fund and counterparty dataset. We assume that these contracts have been terminated and that this was not accurately reported to the trade repositories.

To solve this issue we only want to include contracts with a recent valuation update. To determine the best cutoff we zoom in on the IRS contracts, as these are of interest for the rest of this analysis. We find a similar pattern as for the total portfolio, although less pronounced, as shown in the right panel of Figure 2. We find that while pension funds hold a significant number of contracts that should have been reported as terminated, they also occasionally fail to report valuation updates for contracts that remain open. In contrast, their counterparties, which are mainly large banks, do report recent valuation updates consistently. Therefore, we find that when removing contracts without a recent update, one should rely on the counterparty report when available.

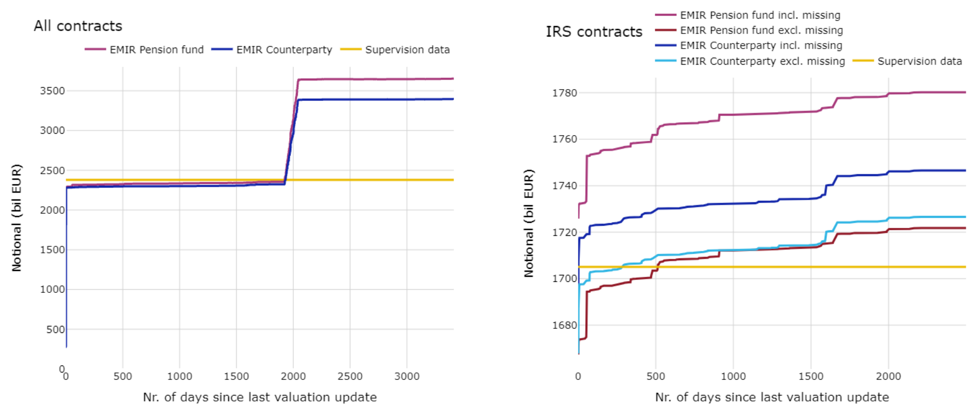


Figure 2 – Total outstanding notional on 29-12-2023 as a function of the maximum included numbers of days since the valuation update. Left panel shows all asset classes, the right panel shows IRS contracts. Sources are FTK data and EMIR data, as reported both by the pension funds and their counterparties. For IRS, we also differentiate in the EMIR data between cases where missing valuation dates are included and those where they are excluded.

In the end we conclude that we find the best possible match by taking the counterparty depulicated data and filtering out all trades with a valution update of more than 5 days in the past. We keep contracts with missing valuation updates, as this gives the best match with supervisory data. We have checked this method not only for the total market, as illustrated in Figure 2, but also for each pension fund individually. We find that this procedure also yields the best alignment between EMIR and FTK data for the greatest number of pension funds. Although we see some outliers for the smaller pension funds, all of the large pension funds match within 2%.

Comparing EMIR data with FTK data not only allows for informed data cleaning but also provides insights into potential biases within the cleaned dataset. In our case, we recognize that removing contracts with outdated valuation dates results in a slight under-reporting in the bilateral sector for certain pension funds. Evaluating results at the pension fund level also enable us to know for each pension fund whether the results are trustworthy or not. For example, when analyzing the characteristics of small pension funds, it seems more appropriate to exclude certain outlier funds.

3 Results

We start this section with a general description of the derivative portfolio of pension funds. The second part of this section analyses the clearing behavior of pension funds over the past few years, in response to the clearing obligation. We also study the availability of clearing members during this time.

The Dutch market consist of 166 pension funds, of which 111 were active in the derivative market in Q4 2023. Table 2 shows the total notional per asset class and contract type outstanding by pension funds at the end of Q4 2023. It confirms that interest rate swaps, along with currency forwards, are the primary derivative products used by pension funds. These products are employed by pension funds to hedge interest rate and currency risk, ensuring stable retirement benefits for participants. We find that in the case of interest rate swaps 94.8% of all contracts are denominated in euro (98.0% based on notional). Given that the clearing obligation applies to EUR IRS we will focus on this contract in the rest of the analysis.

Asset Class/Contract Type	Swap	Future	Forward
Interest Rate	1728	49	
Currency	87		1442
Equity		15	

Table 2 – Total outstanding notional per asset class and contract type held by pension funds as of 29 December 2023. All values are presented in billions of euros. To maintain confidentiality, only categories with ten or more pension funds are displayed.

For this analysis we split the pension funds into four groups, based on their total notional of EUR IRS on 29 December 2023. General statistics of these groups can be found in Table 3. It shows that the largest pension funds, with an outstanding notional of more than €50bn, have the lowest percentage of their portfolio centrally cleared. A potential reason for this could be the lack of clearing members willing to take on their large one-sided positions, as we will discuss below. Although the cleared percentage of the other groups is quite similar, ranging between 53 and 62 percent, we find that in group 4 most pension funds either clear their entire portfolio or not at all. It therefore appears coincidental that the average is of a similar magnitude to that of group 3.

Group	EUR IRS notional range (bil EUR)	Number of pension funds	Above clearing threshold	Percentage of notional cleared	Average maturity (years)
1	≥ 50	6	Y	45.2	22.4
2	10 – 50	11	Y	52.8	28.3
3	3-10	18	Y	62.0	26.8
4	≤ 3	76	N	62.0	26.5

Table 3 – Pension funds characteristics based on EUR IRS on 29 December 2023.

Figure 3 shows the evolution of the clearing rate for each of these groups. It shows a steady increase in clearing for all groups towards the introduction of mandatory central clearing at the end of Q2 2023, indicating that pension funds tried to prepare well in advance. Furthermore, it indicates that the largest pension funds in group 1 clear a significantly smaller amount than their smaller counterparts, even though their clearing rate increased from 18% to 45% during this period. We anticipate that this will change in the future as the bilateral contracts of pension funds exceeding the clearing threshold are replaced by cleared transactions. However, due to the long average maturity of IRS, we expect pension funds to engage in transactions infrequently, meaning that this transition will take several years.

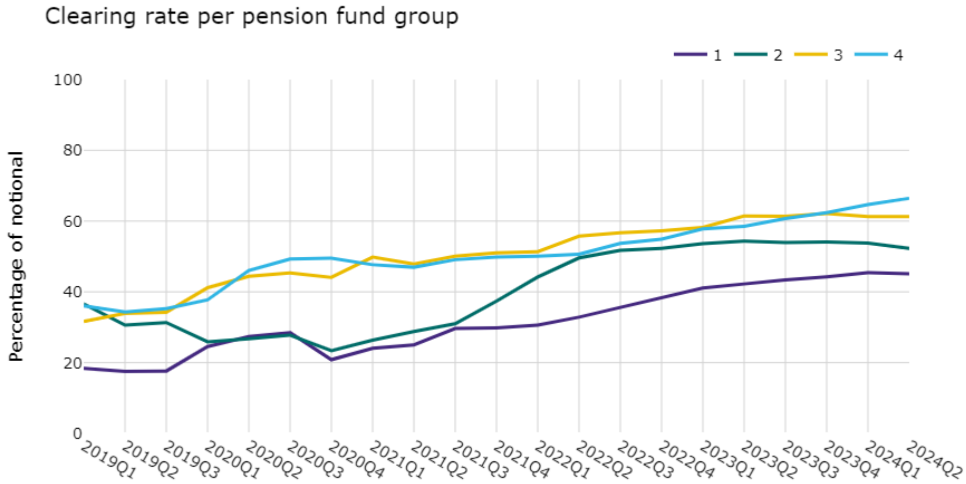


Figure 3 – Clearing rate per pension fund group based on outstanding notional. The fraction is calculated as the average per pension fund per day.

The limited availability of potential clearing members in the past may be the reason for the low clearing rates among large pension funds. Figure 4 illustrates the market concentration among the major clearing members throughout the entire period. It shows that at the start of 2019 the top 3's clearing members held 83% of the market share. Recently, the top 3 have dropped to 63%, indicating a larger availability of potential clearing members. This is probably due to the introduction of obligatory clearing in combination with the introduction of margin requirements for bilateral contracts.

Furthermore, we find that a pension fund's current clearing members are primarily the same counterparties that maintained bilateral relationships with the fund before the clearing obligation. Rather than establishing new relationships with large international clearing members, most clearing activity occurs through these existing connections.

4 Conclusion and next steps

We have analyzed the effects of the central clearing obligation on Dutch pension funds; the obligation came into effect in June 2023. In our analysis, we focused on EUR IRS because they are subject to mandatory central clearing and are commonly used by pension funds to hedge their interest rate risk. We find that the clearing rate has steadily increased since 2021 in preparation for mandatory clearing. The clearing rates of larger pension funds lag behind those of smaller funds, due to their

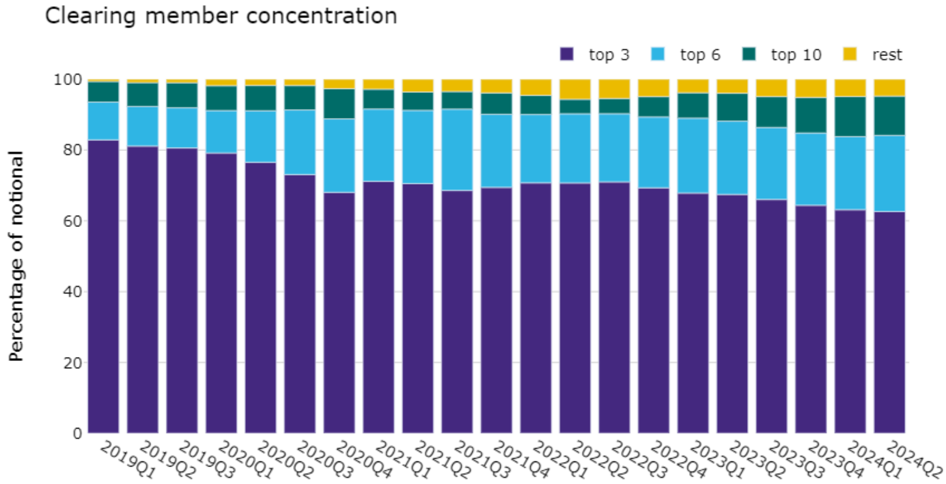


Figure 4 – Market concentration for the major clearing members as percentage of the total notional. Each category is based on the daily average of that quarter.

large one-sided positions. In contrast, pension funds below the clearing threshold either clear their entire portfolio or not at all. Furthermore, we find that the clearing obligation has probably resulted in the fact that more counterparties have become available as potential clearing members for pension funds, leading to a less concentrated market.

As the newly available clearing members are the same as the original bilateral counterparties of pension funds, an interesting open question is who takes on the other sides of these cleared contracts and whether these are the same counterparties. We hope to answer this question in the near future, since we are now able to analyze the whole euro area dataset thanks to EMIR Refit.

Bibliography

DNB (2023): "Register of pension funds," https://www.dnb.nl/en/public-register/register-of-pension-funds/?p=2&l=100&rc=UFdQTkY.

ESMA (2019): "Clearing thresholds," https://www.esma.europa.eu/post-trading/clearing-thresholds.

European Union (2012): "European Market Infrastructure Regulation," https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32012R0648.

Perez-Duarte, S. and G. Skrzypczynski (2019): "Two is company, three's a crowd: automated pairing and matching of two-sided reporting in EMIR derivatives' data," in *IFC conference "Are post-crisis statistical initiatives completed?"*, Bank for International Settlements, no. 49 in IFC Bulletin, https://www.bis.org/ifc/publ/ifcb49_51.pdf.

Derivatives of Belgian banks and insurance companies: Three case studies into exposures, rising interest rates, and the energy crisis

Saif Ben Hadj,[a,1] Riemer Faber,[a]
Pablo Rovira Kaltwasser,[b] and Patrick Van Roy[a,c]

[a] *Financial Stability, AML Supervision and Banking Prudential Policy, National Bank of Belgium, Brussels, Belgium*
[b] *General Statistics, National Bank of Belgium, Brussels, Belgium*
[c] *Solvay Brussels School of Economics and Management, Université Libre de Bruxelles, Brussels, Belgium*

Abstract. This paper discusses three examples of how the National Bank of Belgium uses derivatives data for financial stability analysis in Belgium: (1) the monitoring of derivatives exposure of banks and insurance companies; (2) an assessment of the impact of rising interest rates on interest rate derivatives; and (3) a study of the role of commodity derivatives during the energy crisis. Among others, it finds that banks cleared a substantial share of their interest rate derivatives outside the European Union during the period 2021-2023. Furthermore, in line with their business models, the market value of banks' interest rate derivatives in general increased after the 2022-2023 rise in interest rates, while the market value of interest rate derivatives of insurance companies in general declined. Commodity derivatives were a moderate risk for financial stability during the energy crisis.

[1] This contribution is a shortened version of a thematic article from the 2024 Financial Stability Report of the National Bank of Belgium (*Derivatives through the lens of EMIR and supervisory data*, pp. 133-148). The authors are grateful to participants in the Banca d'Italia-ESRB workshop *EMIR data analytics for research, financial stability and supervision*, and various seminars at the National Bank of Belgium for their helpful suggestions. The views expressed are those of the authors and do not necessarily reflect those of the National Bank of Belgium.

The direct exposure of the financial sector was relatively limited, and the indirect exposure of banks via credit risks appeared to be manageable. However, for some non-financial firms the large movements in commodity prices had a large impact on the outcomes of their derivatives contracts.

Key words: Derivatives, EMIR, Belgium, central clearing, interest rates, energy crisis.

1 Introduction

Past financial crises have demonstrated that derivatives can amplify or even trigger financial instability. Therefore, regulation and high-quality data on derivatives trading are essential to ensure effective, evidence-based supervision of derivatives and thus financial stability. The European Market Infrastructure Regulation (EMIR) is the main regulation on derivatives in the European Union (EU) and it plays an important role in achieving these objectives.

The National Bank of Belgium (NBB) uses data from EMIR reporting as well as derivatives data from other sources, such as supervisory reporting data, for monitoring financial stability in Belgium. This contribution discusses three examples: (1) the monitoring of the derivatives exposure of banks and insurance companies; (2) an assessment of the impact of rising interest rates on interest rate derivatives; and (3) a study of the role of commodity derivatives during the energy crisis.

First, Section 2 reports descriptive statistics on the exposure of Belgian banks and insurance companies to derivatives. It shows that the notional amount is much larger for banks than for insurance companies, although both banks and insurance companies have a substantial exposure to derivatives, in particular interest rate derivatives with a long maturity. Furthermore, during the period 2021-2023, banks in particular cleared a substantial share of their interest rate derivatives outside the EU. In that regard, EMIR has recently been updated to improve the attractiveness of EU central counterparties and reduce the reliance on systemically important third-country central counterparties.

Second, Section 3 assesses the impact of the 2022-2023 rise in interest rates on the interest rate derivatives held by Belgian banks and insurance companies. Banks tend to have interest rate derivatives in which they pay a fixed amount to the counterparty and receive a floating interest rate in return (for example, to hedge households' fixed interest mortgage loans). In general, the market value of banks' interest rate derivatives substantially increased after the rise in interest rates. The value of their posted collateral fell. Insurance companies tend to have interest rate derivatives in which they pay a floating interest rate and receive a fixed amount (as in general the revenues on their assets vary with interest rates while their payouts do not). In general, the market value of their interest rate derivatives decreased after the rise in interest rates. However, as insurance companies use interest rate derivatives less extensively than banks, the impact on the value of the collateral they had to post was relatively limited.

Third, Section 4 studies the role of commodity derivatives during the energy crisis. Commodity derivatives were a moderate risk for financial stability in Belgium during the period of high and volatile commodity prices that started after the Covid-19 pandemic and intensified following Russia's invasion of Ukraine. The direct exposure of Belgian banks and insurers was relatively limited as the notional value of the commodity derivatives they held was relatively modest. The indirect exposure of banks via credit risks appeared to be manageable. Commodity derivatives had various characteristics that, in addition to their volume, impacted the indirect exposure of banks. For example, a large share of the commodity derivatives were held by a small number of non-financial firms, and many commodity derivatives had a relatively short residual maturity. Despite the moderate risks for financial stability, for some non-financial firms the high and volatile commodity prices had a substantial impact on the outcomes of their derivatives contracts.

These three examples confirm the policy relevance of the EMIR data. For example, the high frequency of the EMIR data was important when commodity prices and market values of commodity derivatives changed rapidly during the energy crisis. The granularity of the EMIR data revealed important characteristics of the derivatives contracts such as the country in which derivatives were cleared and the maturity dates. The EMIR data are thus a helpful source of information for financial stability analysis.

2 Derivatives exposure of banks and insurance companies

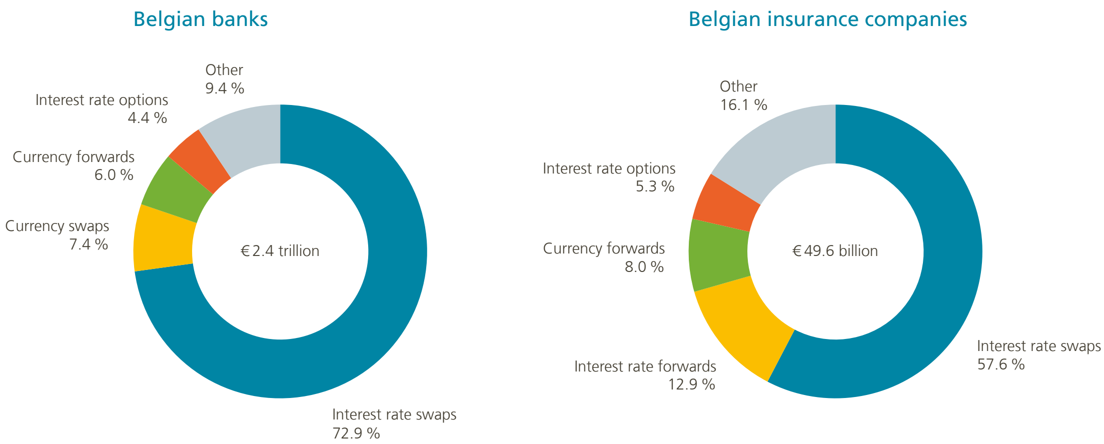
EMIR data are used to compile descriptive statistics on the exposure of Belgian banks and insurance companies to derivatives.

At the end of December 2023, Belgian banks had significantly higher notional derivatives exposure (€2.4 trillion) than insurance companies (€49.6 billion).[2] Two main factors explain this finding. First, banks use derivatives to hedge maturity gaps in their (loan) portfolios; this is an intrinsic characteristic of the banking business (see the next section). Second, banks often act as intermediaries when providing financial services, in particular to non-financial corporations that wish to hedge their risk, by serving as the derivative counterparty while simultaneously hedging this position with another financial institution (a so-called back-to-back hedge).

Figure 1 provides a breakdown of the most common derivatives used by Belgian banks and insurance companies. Interest rate derivatives (swaps, forwards and options) are the most commonly used derivatives by Belgian banks (77.3%) and insurers (75.9%). As can be seen, interest rate swaps represent the largest share by far (72.9% for banks and 57.6% for insurance companies), followed by currency swaps (7.4%) and currency forwards (6.0%), in the case of banks, and by interest rate forwards (12.9%) and currency forwards (8.0%), in the case of insurance companies.

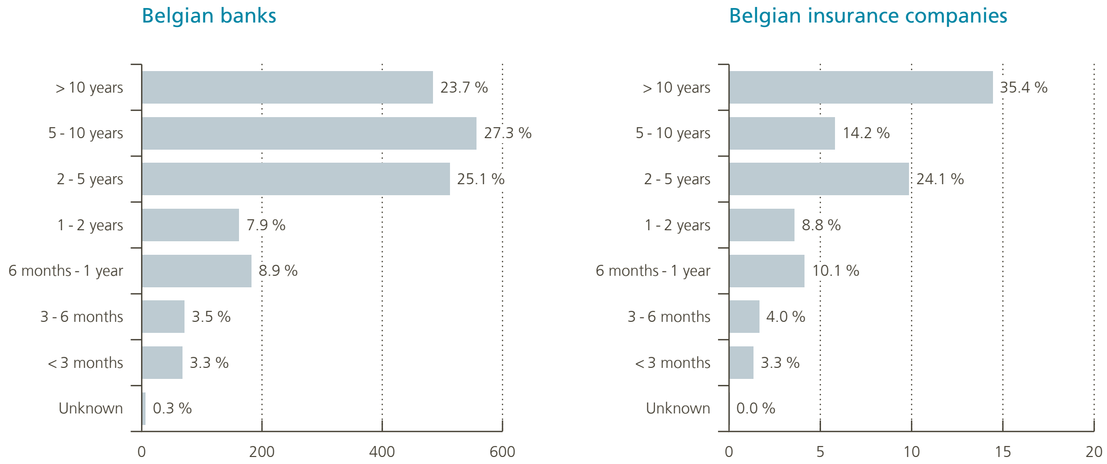
Since interest rate derivatives are the most commonly used by banks and insurance companies in Belgium, the figures presented in the remainder of this section

[2] Total notional values are calculated as the sum of the entity-specific notional values (and hence include the same contract twice if two of the entities trade with each other).



Notes: %, end of 2023.
Source: NBB (EMIR reporting).

Figure 1 – Notional amounts of derivatives held by Belgian banks and insurance companies, broken down by type of derivative.



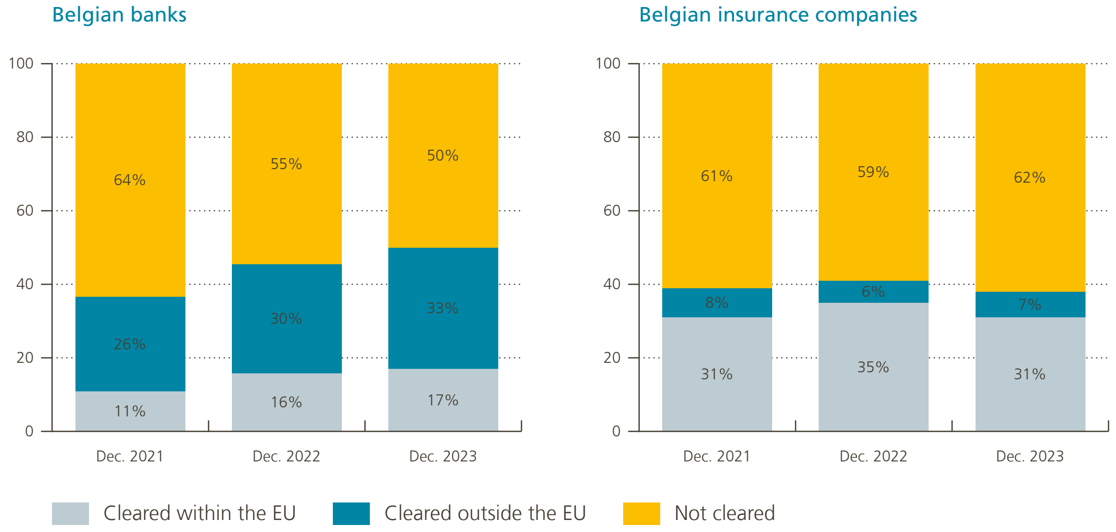
Notes: Notional values, in € billion and in % of total, end of 2023.
Source: NBB (EMIR reporting).

Figure 2 – Breakdown of the interest rate derivatives of Belgian banks and insurance companies by original maturity bucket.

focus exclusively on this type of derivative.

Most interest rate derivatives of Belgian banks and insurers tend to have a relatively long maturity. As can be seen from Figure 2, 51% of banks' interest rate derivatives have an original maturity of longer than five years. This figure is quite similar for insurance companies, at approximately 50%. Interest rate derivatives with an original maturity of between one and five years make up approximately 33% of the interest rate derivative portfolios of banks and insurers. Only 3.3% of their interest rate derivatives have an original maturity of three months or less.

One of the main requirements of EMIR is the obligation to clear the most standardised, and therefore most liquid, interest rate derivatives and credit derivatives with a central clearing counterparty. Figure 3 illustrates the share of outstand-



Notes: % of notional value, end of 2023.
Source: NBB (EMIR reporting).

Figure 3 – Breakdown of the interest rate derivatives of Belgian banks and insurance companies by clearing status.

ing interest rate derivatives centrally cleared by banks and insurance companies in Belgium. At the end of 2023, approximately 50% of the outstanding interest rate derivatives of banks was centrally cleared; for insurance companies this was approximately 40%.

Two additional observations can be drawn from this figure. First, and in line with the policy objective of achieving a high degree of central clearing, the share of bank interest rate derivatives that are cleared is increasing (+14 percentage points between the end of 2021 and 2023). A substantial proportion of Belgian bank interest rate derivatives that were not cleared at the end of 2023 relates to intragroup transactions, the majority of which benefit from an exemption from the central clearing obligation granted by the supervisor (in this case, the NBB), a possibility that is provided for by EMIR if certain conditions are met.[3] Second, banks clear more interest rate derivatives outside the EU (33%) than within (17%), whereas the reverse holds true for insurers (7% versus 31%). However, when considering the absolute figures for both sectors combined, the majority of interest rate derivative clearing still occurs outside the EU. The 2024 update of EMIR aims to impose stricter regulatory requirements on the location of clearing activities, potentially centralising these activities within the EU to enhance financial stability and oversight.

[3] These conditions require among other things that risk management procedures be adequately sound, robust and consistent with the level of complexity of the transaction and that there be no impediment to the prompt transfer of own funds or the repayment of liabilities between the counterparties (see Regulation (EU) No 648/2012).

3 Impact of rising interest rates on the market value of derivatives

This section explores the impact of the 2022-2023 rise in interest rates on both the market value of interest rate derivatives traded by banks and insurance companies and the amount of collateral they were required to post. As EMIR reporting is still relatively new, this analysis is based on supervisory reporting data which cover a longer time span. It is prefaced by a brief discussion of why and how banks and insurance companies use interest rate derivatives for risk management purposes.

Belgian banks tend to have a rather high exposure to interest rate risk. On the assets side, they hold a relatively large volume of assets whose interest rates are fixed for a long period of time, such as mortgage loans. On the liabilities side, these assets are mainly financed by (retail) sight and savings deposits with no contractual maturity or repricing date (so-called "non-maturity deposits"). As a result, Belgian banks tend to have a relatively wide maturity gap between their assets and liabilities which makes them vulnerable to repricing risk arising from timing differences in changes in interest rates for instruments on both sides of the balance sheet. As shown in the previous section, banks hedge this risk using a substantial volume of interest rate derivatives.

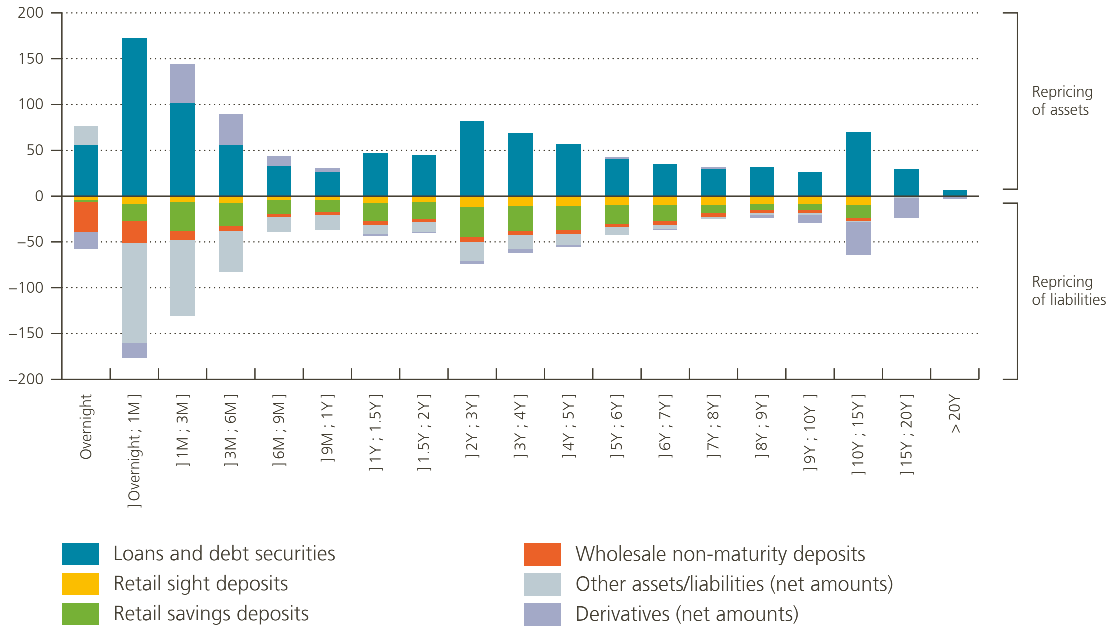
The use of interest rate derivatives to hedge interest rate risk is further illustrated in Figure 4. This chart reveals how the Belgian banking sector manages interest rate risk in the banking book, showing the notional amounts of assets, liabilities and derivatives according to their remaining term to maturity, at which time the interest rate for the position will be adjusted to market conditions. In the case of positions with an interest rate fixed for the entire term, the notional amount is placed in the repricing bucket when the position reaches maturity. The chart shows that banks try to keep the net gap between the asset and liability positions in the different repricing buckets at a low level, so as to reduce (net) exposure to interest rate changes. Derivatives are also used to reduce these gaps, notably in longer-term buckets where the interest rate risk on fixed-rate mortgages is mitigated by payer swaps.[4]

These interest rate derivatives are managed in a dynamic way by banks – using payer and receiver swaps, the latter also being used to close previous payer swap positions – but, overall, they result in a net payer swap position for the banking system, whereby Belgian banks pay interest at a fixed long-term rate to the derivative counterparty and receive interest at a floating short-term rate in return. Of course, when interest rates rise, the floating-rate leg of these interest rate derivatives will generate higher income, while the fixed-rate leg remains at the same level until the end of the contract. The upper charts in Figure 5 confirm that hedging by Belgian banks against a rise in interest rates paid off in the form of a significant net increase in the market value of their interest rate derivatives up to the third quarter of 2023.[5]

---

4 A payer swap entails paying a fixed amount and receiving variable interest. A receiver swap entails paying variable interest and receiving a fixed amount. Hence, a swap agreement between two parties is a payer swap for one and a receiver swap for the other.

5 The correlation between the level of long-term (10-year) interest rates and the net market



Notes: Consolidated data for the six largest banks, after considering modelling assumptions, September 2023, in € billion.
Source: NBB (STE IRRBB reporting).

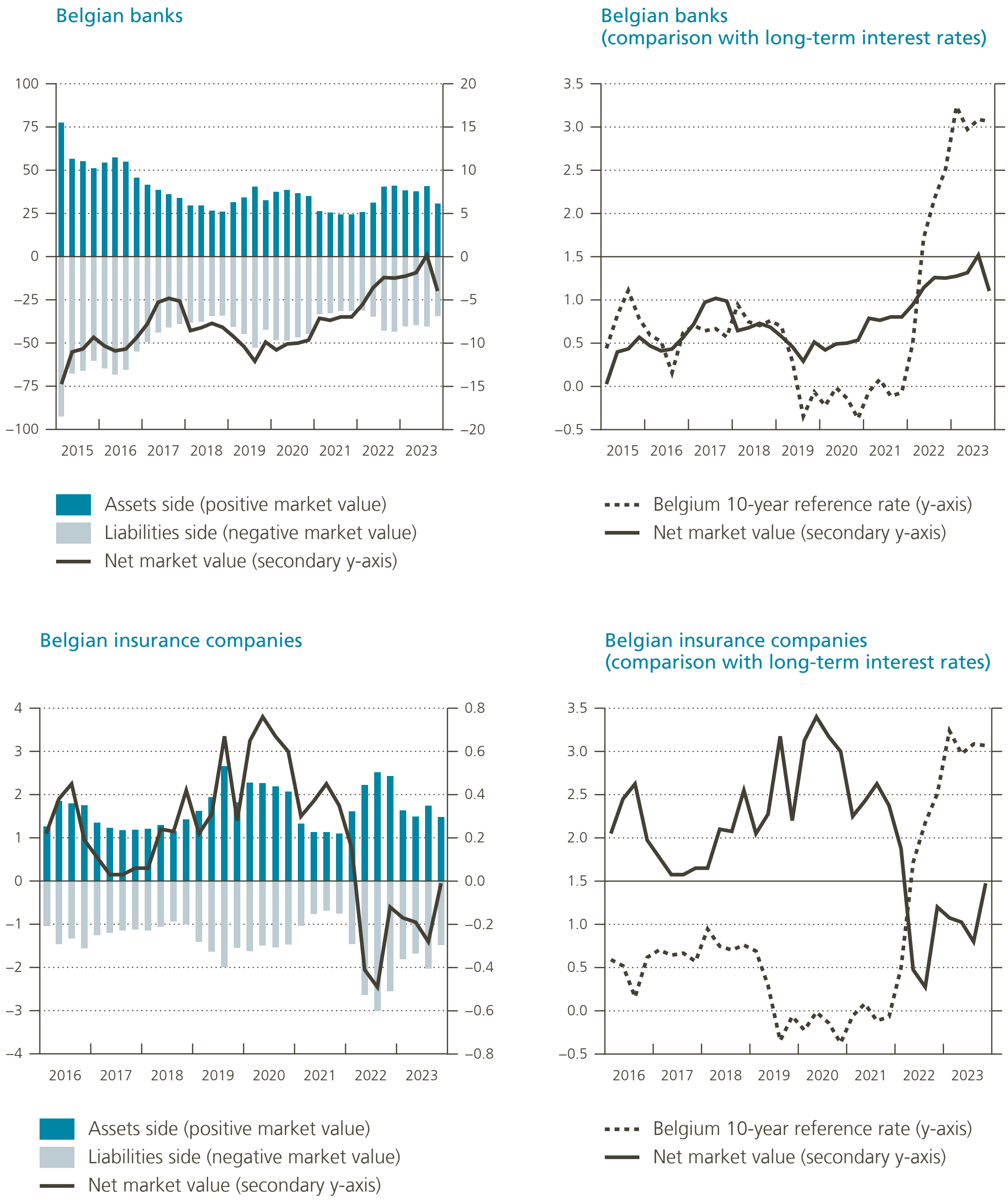
Figure 4 – Maturity schedule for notional repricing cash flow of assets and liabilities in the banking book.

Unlike for banks, the liabilities of insurance companies usually have longer maturities than their assets, as liabilities are based on long-term contracts. Derivatives, in particular interest rate swaps, are an important tool used by insurers to mitigate the risks inherent in their long-term business. The use of interest rate swaps by insurance companies typically results in a net receiver swap position, whereby they pay interest at a floating rate to the derivative counterparty and receive interest at a fixed rate in return. While these swap contracts benefit from a fall in interest rates, the opposite is true when interest rates rise. The lower charts in Figure 5 confirm that, following the rise in interest rates, Belgian insurance companies saw a significant net decrease in the market value of their interest rate derivatives.[6]

Higher interest rates affect the market value of interest rate derivatives and consequently may lead to changes in the collateral that needs to be posted with the counterparty holding the positive market value of the contract. Generally speaking, the Belgian banking sector is likely to experience a positive net liquidity inflow when interest rates rise, as it will have to post less collateral with interest rate derivative counterparties than before. In contrast, insurance companies are generally on the opposite side due to the lengthy maturities of their liabilities and corresponding net receiver swap position, which helps close the gap with the duration of their assets. However, since the insurance sector uses interest rate derivatives less extensively than the banking sector, the impact of higher inter-

value of interest rate derivatives is strongly positive (72.7%) for banks.

[6] The correlation between the level of long-term (10-year) interest rates and the net market value of interest rate derivatives is strongly negative (-83.6%) for insurance companies.

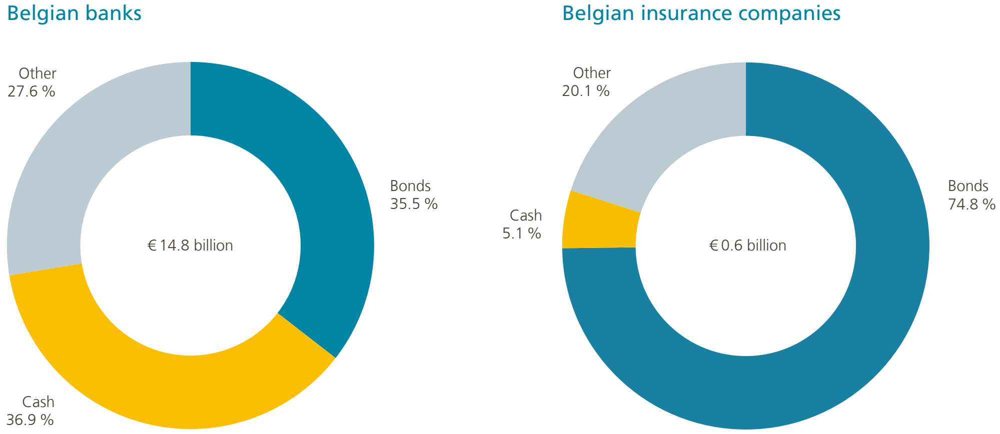


Notes: Market value in € billion and long-term interest rates in %, consolidated data.
Source: NBB (Finrep and QRT reporting).

Figure 5 – Market value of interest rate derivatives held by Belgian banks and insurance companies and level of long-term interest rates.

est rates on the value of the collateral that insurance companies need to post is relatively limited.

Figure 6 indeed confirms that, by the end of 2023, the value of the collateral posted by insurance companies for derivative transactions (€0.6 billion) was significantly lower than that posted by Belgian banks (€14.8 billion). Although substantial, the value of collateral posted by banks has fallen from the period before the interest rate rises (€21.8 billion at the end of 2021). Figure 6 further shows



Notes: %, end of 2023.
Source: NBB (Finrep and QRT reporting).

Figure 6 – Total collateral provided by Belgian banks and insurance companies for derivative transactions, broken down by type.

that the most common types of collateral posted by Belgian banks and insurance companies are liquid assets (cash and bonds).[7]

In sum, although large margin payments could theoretically pose financial risks, Belgian banks have had to post less collateral in the wake of the 2022-2023 interest rate rises, while the collateral posted by Belgian insurance companies has remained limited.

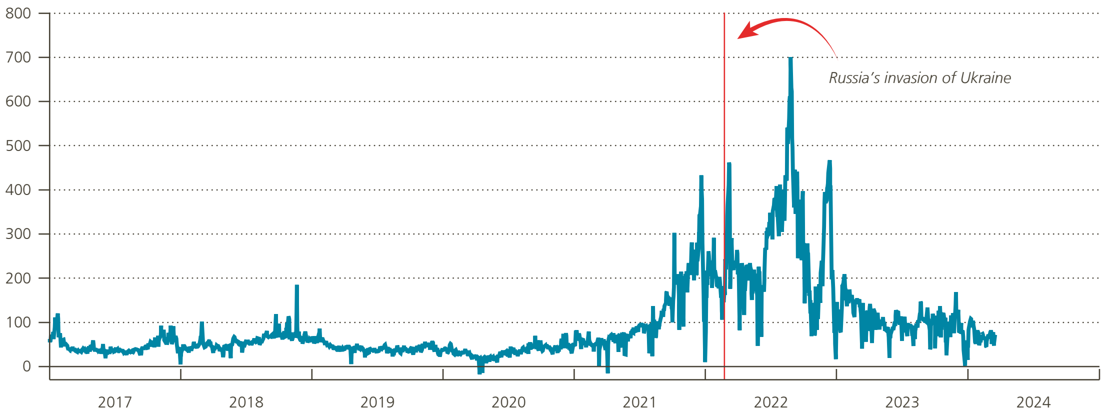
4 Commodity derivatives and the energy crisis

This section focuses on the exposure of the Belgian financial sector to commodity derivatives, both directly (through the holding of such derivatives) and indirectly (through their exposure to counterparties holding such derivatives).

Commodity prices increased and became more volatile in the aftermath of the Covid-19 pandemic. Russia's invasion of Ukraine aggravated these developments and resulted in a period of extremely high and volatile commodity prices. A visible consequence of this was the global energy crisis that followed the surge in, among others, electricity and gas prices; however the prices of, for example, wheat and metals (nickel) also experienced strong fluctuations. To illustrate these extreme price movements, Figure 7 shows the daily spot price of electricity. At its peak in August 2022, the price was 20 times higher than at the start of 2020. More recently, prices for electricity and many other commodities have returned to the levels seen before the start of the war in Ukraine.

Derivatives are used to trade commodities and hedge commodity prices. For example, an oil company can use a futures contract to sell a barrel of oil at a given price at a specific point in time in the future (the notional value of the contract is the current market price of the barrel). Similarly, an electricity retailer that sells

[7] For insurance companies, the breakdown of collateral provided for derivatives is extrapolated based on the total breakdown of collateral.



Notes: In € per megawatt-hour.
Source: Refinitiv.

Figure 7 – Daily spot prices of electricity.

consumers electricity at a fixed retail price but purchases electricity at the spot price can insure itself against future cost increases by entering into a swap contract in which, at specific times in the future, it pays a fixed amount and receives the spot price for a megawatt-hour of electricity.

When commodity prices are in turmoil, the market value of a derivatives contract (that is, the sum of the present values of all amounts a party expects to pay and receive in the future) can suddenly change substantially. For example, after Russia's invasion of Ukraine and the subsequent increase in electricity prices, the abovementioned swap contract would be substantially more valuable for the electricity retailer and substantially more costly for its counterparty. Basically, the invasion caused an unexpected loss for the electricity retailer's counterparty. It should be noted, of course, that this counterparty could be exposed to electricity prices in other ways besides just this contract and that the overall impact of the invasion on this counterparty could thus differ.

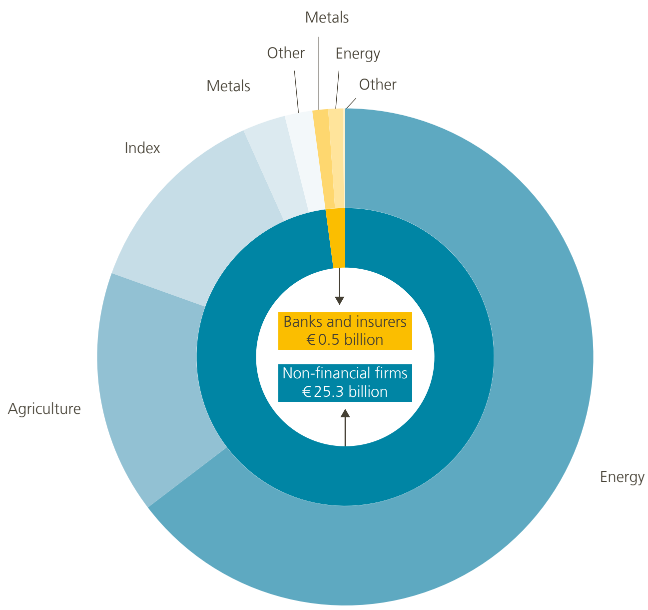
Furthermore, strong swings in commodity prices affect the collateral a party provides to ensure it can fulfil its obligations should it default. Contracts regularly provide for the exchange of variation margin and initial margin. The variation margin typically covers the market value of the contract, and changes in the value are exchanged daily. The initial margin covers the costs of default arising in the period between the final payment of variation margin and the liquidation or hedging of the contract. In general, higher and more volatile commodity prices lead to more extreme collateral requirements which are more difficult to predict. Usually, a party must provide variation margin in cash at short notice, which can give rise to liquidity risk.

Entities with large losses on commodity derivatives and considerable short-term collateral requirements can pose a threat to financial stability. Therefore, in particular in the period since the turmoil on commodity markets, the NBB closely monitors commodity derivatives to prevent risks on commodity markets (in particular, risks related to the energy sector) from spilling over to the financial sector. The EMIR dataset is an important source of information for these analyses as it contains daily information on individual commodity derivatives.

4.1 Direct exposure

The direct exposure of the Belgian financial sector to commodity derivatives was relatively limited during the energy crisis. For example, in February 2023, about a year after Russia's invasion of Ukraine, the total notional value of commodity derivatives reported by Belgian parties was €25.9 billion. Only €0.5 billion (2%) of this amount was in the hands of banks and insurers (see Figure 8). By comparison, the total value of the assets of banks and insurers is approximately €1 500 billion. This shows that Belgian banks engage in a limited amount of trading and hedging on their own behalf, but also that their role as an intermediary to assist clients is limited. In fact, Belgian non-financial firms with a substantial amount of commodity derivatives tend to have contracts with foreign banks (if they have contracts with a bank). Energy products account for somewhat less than half of the notional value of the commodity derivatives held by banks and insurers (see Figure 8).[8]

During the period of turbulence on the commodity markets, the NBB regularly used these and other statistics to monitor the direct exposure of banks and insurers to commodity derivatives.



Notes: Data for 7 February 2023.
Source: NBB (EMIR reporting).

Figure 8 – Notional value of commodity derivatives (€25.9 billion) by reporting sector (inner ring) and commodity type (outer ring).

8 The exposure of financial and non-financial firms to commodity prices may be larger than the reported figures as derivatives are not the only way in which entities can be exposed to commodity prices (for example, they can also own commodities directly).

4.2 Indirect exposure

Banks may be indirectly exposed to commodity derivatives if their clients are not able to pay back loans due to losses on commodity derivatives or liquidity shortages. Non-financial firms hold €25.3 billion in commodity derivatives.[9] These firms are active in various sectors, including energy, food (both production and trading), transport and industrial inputs. About two-thirds of the commodity derivatives owned by non-financial firms relate to energy products. The remaining one-third are for agricultural products and metals, among others (see Figure 8). A large share of commodity derivatives are held by a small number of non-financial firms. It appears that some non-financial firms actively changed their derivative positions during the crisis. For example, some strongly increased the number of contracts over time.

Indirect exposure to commodity derivatives appeared manageable for banks during the energy crisis, also because the total notional amount was relatively small. Furthermore, supervisory data show that the loan exposure of Belgian banks to the energy sector has been moderate and stable over time. In the first quarter of 2023, 4% of all loans to non-financial firms were to those operating in the Belgian and foreign energy sector.

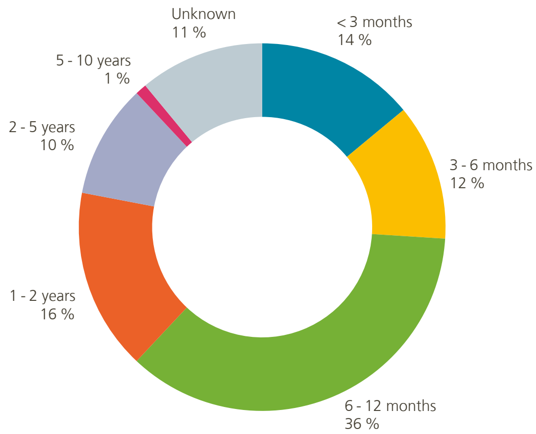
In addition to the volume of commodity derivatives, various other factors impact the indirect exposure of banks, including the following. First, highly concentrated ownership of commodity derivatives could give rise to a situation whereby a small number of non-financial firms bear a large proportion of the total losses on commodity derivatives and default on their bank loans; instead of, for example, a situation where a large number of firms all sustain a relatively small loss and are able to pay back their loans.

Second, a large share of commodity derivatives are traded over the counter rather than on an exchange. Over-the-counter derivatives may be subject to less strict collateral requirements and thus give rise to larger losses for a non-financial firm should its counterparty default. These losses could impact the firm's ability to meet its bank loan payments.

Third, a substantial share of Belgian commodity derivatives are contracts between Belgian firms and foreign firms belonging to the same conglomerate (i.e., intragroup contracts). A foreign firm may serve as an intermediary and pass a contract on to a foreign bank, for example, but it is also possible for the foreign firm to act as the ultimate counterparty in the transaction. In any case, conglomerates may constitute a first line of defence in the event of losses, thereby possibly reducing the likelihood of firms defaulting on their bank loans. However, intragroup contracts may be backed by less collateral than contracts between firms that do not belong to the same conglomerate, which would increase the risk for banks.

Fourth and finally, shortly after Russia's invasion of Ukraine in February 2022, over 60% of commodity derivatives had a residual maturity of up to a year (see Figure 9). Contracts with a remaining duration of more than two years were uncommon, and contracts with a remaining duration of more than five years were

[9] The Belgian Financial Services and Markets Authority (FSMA) can grant to non-financial firms an exemption for reporting intragroup contracts. Therefore, the reported value is a minimum value.



Notes: Data for 1 April 2022, contracts are weighted by notional value.
Source: NBB (EMIR reporting).

Figure 9 – Residual maturity of commodity derivatives (shortly after Russia's invasion of Ukraine).

rare. Hence, at the end of 2023, contracts that started before or just after the invasion mostly had ended and been replaced with new contracts for which the start of the war was known information. Most of the issues caused by commodity derivatives outstanding around the time of the invasion should have materialised by then.

Despite moderate risks for financial stability, the large movements in commodity prices had a substantial impact on the outcomes of derivatives contracts for some non-financial firms. An in-depth look at the data of some firms shows that the market value of their derivatives contracts moved closely in line with the prices of the contract commodities. Depending on their position, for some firms the market value greatly increased following Russia's invasion of Ukraine, while for others the market value fell substantially. These extreme market valuations faded away when commodity prices returned to their pre-war levels and contract payments returned to their old values. Another contributor to less extreme market values was that, over time, existing contracts came closer to maturity (under the abovementioned swap contract, for example, fewer future payments were left) or reached maturity and were replaced with new contracts based on the latest information. Similarly, firm-level data show that some firms had to post – while others received – large amounts of collateral after Russia's invasion of Ukraine (both initial margin and variation margin). The variation margins closely followed market values, meaning some firms provided large amounts of cash at peaks in the market value. Variation margin payments declined along with commodity prices, so that firms no longer had to provide large amounts of cash to meet collateral requirements and liquidity risks subsided substantially.

Luxembourg derivatives market 2023 CSSF & CAA joint report on market structure, market trends and data quality

Andrea Gentilini[a]

[a]*Financial Market Infrastructure Division, Commission de Surveillance du Secteur Financier, Luxembourg, Luxembourg*

Abstract. The Commission de Surveillance du Secteur Financier (CSSF) and the Commissariat aux Assurances (CAA) published on 11 March 2024 the first joint report on the Luxembourg derivatives market. The report aims to provide an overview of the state of play under EMIR, while also providing insights into ongoing efforts to improve the quality of the data. This paper summarizes the main findings and methodological approaches of the joint report, emphasizing its relevance for financial stability monitoring and supervision.
Key words: EMIR, CSSF & CAA joint report.

1 Ojective of the CSSF & CAA joint report

The Commission de Surveillance du Secteur Financier (CSSF) and the Commissariat aux Assurances (CAA) published on 11 March 2024[1] the first joint report on the Luxembourg derivatives market. The report highlights the activity in derivative instruments by counterparties established in Luxembourg as well as the supervisory activities carried out by CSSF and CAA regarding the quality of data reported to Trade Repositories (TRs) under the European Market Infrastructure Regulation (EMIR).[2] This report aims to provide an overview of the state of play

[1] See CSSF and CAA [2023b].
[2] 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.

under EMIR, while also providing insights into CSSF and CAA's ongoing efforts to improve the quality of the data (see CSSF [2022] and CSSF and CAA [2023a]).

The first part of the report describes the Luxembourg derivatives market structure and its trends; it also provides key information on the type of counterparties active in the derivatives market, the asset classes and other relevant high-level information. In addition, the quantitative analysis in the first part of the report underscores the existing scope for enhancing the data quality in EMIR reporting at EU and Luxembourg levels. In this context, CSSF and CAA encourage all stakeholders, and in particular counterparties involved in derivatives trading activity, to use and leverage the regulatory data in their internal processes, such as – but not limited to – their risk and compliance management processes.

In the second part, the report provides a description of the data quality engagement framework (DQEF),[3] the related indicators and the work performed by ESMA and national competent authorities (NCAs) as well as an overview of the Luxembourg reported data.

2 Focus on the data quality

2.1 Importance of data quality assessment

Transaction-level data play a key role in the daily operations of NCAs and ESMA. More than 50 authorities and institutions at EU level and at national level of each Member State receive EMIR data. Each of these authorities/institutions has its own specific mandate and responsibilities. To fulfil their mandate related to the derivatives market, these authorities/institutions extensively use EMIR data, notably for the purposes related to financial market stability, maintaining orderly markets, and ensuring market integrity. Furthermore, regulatory data can and shall be used by supervised entities for their own internal purposes such as reporting to the board, risk management activities, and reporting to their clients. Indeed, the extensive use of regulatory data by supervised entities, as affirmed on several occasions by both CSSF and CAA, will improve the quality and reliability of the data, reducing the regulatory burden and increasing the level of compliance. CSSF and CAA echo the considerations made by the ESRB [2022b] according to which poor data quality: (i) impedes the adequate monitoring of (financial stability) risks by authorities, which was one of the goals of the post-crisis reforms; (ii) compels policymakers to devote substantial time and resources to follow up on data quality; (iii) creates blind spots due to the exclusion from monitoring of entities reporting implausible values; and (iv) may be symptomatic of a more fundamental problem of poor risk management among certain reporting entities.[4]

From supervisory experience, data quality is only the visible part of larger issues that occur within the entity or the network of entities in the reporting process. Indeed, sound and robust IT systems, controls and processes, data models and

3 ESMA DQEF has been established at the European level to monitor data quality on an ongoing basis and to ensure adequate supervisory engagement with the supervised entities, namely TRs and reporting counterparties.

4 See ESRB [2022a] and ESRB [2022b].

governance structures must be in place to ensure high data quality. Nevertheless, most entities manage EMIR obligations in a silo-approach where processes are not fully under control and the EMIR data are not well organised.

2.2 Derivative market in Luxembourg

At the Luxembourg level, at the end of June 2023, trade repositories reported a total of slightly more than 914,000 open transactions amounting to a gross notional outstanding amount of around EUR 6,484bn, including both over the counter (OTC) and exchange traded (ETD) derivatives. In notional terms, at the end of June 2023, currency derivatives and interest rate derivatives dominated the market, with 42% and 35% respectively of the total amount outstanding. Looking at the number of outstanding derivative contracts, currency derivatives represent almost 50% of the total, followed by equity derivatives (34%) and interest rates derivatives (11%).

Overall, forwards, swaps and futures are the main contract types used on the Luxembourg market. However, swaps are the most used contract type for credit derivatives (85%) while forwards are the most used derivative contract for currency derivatives. Concerning the remaining maturity of derivatives, short-term maturities prevail in terms of notional with more than 70% of the derivatives having less than one-year remaining maturity. Short term maturities prevail in currency, commodity, and equity derivatives too, while for credit derivatives maturities above 1 year represent the largest component.

Investment funds are the main participants in derivatives market, accounting for more than 62% of trading activity in both notional terms and number of outstanding transactions; credit institutions represent slightly more than 10% in notional terms and about 17% in number of outstanding transactions. OTC derivatives significantly prevail with more than 80% of the notional compared to ETD, with ETD mainly used in interest rate, commodity, and equity derivatives, but less than 1% for the remaining asset classes. With regards to cleared rates, interest rate derivatives are by far the most cleared asset class, representing more than 70% of all cleared notional, followed by equity and credit derivatives with less than 20% each. Euro (EUR) and US dollar (USD) are the most relevant currencies used to report notional values.

The Luxembourg derivatives market is not very concentrated. Indeed, since 905 counterparties represent 80% of the total notional amount. However, looking at the concentration per asset class, the levels vary considerably. For currency derivatives, slightly more than 1,000 counterparties represent 80% of the market. For commodity derivatives only 36 counterparties cover 80% of the total notional.

2.3 Statistical methods and data preparation

The statistics presented in the report are based on the reporting requirement specified in EMIR and its related technical standards adopted for its implementation. With reference to the statistical standards and methods, all derivative transactions reported by or on behalf of a counterparty established in Luxembourg under the supervision of the CSSF or the CAA have been considered. This includes all derivatives instruments, currencies, maturities, and trading venues as reported by entities. Similar to the equivalent analysis performed by ESMA [2023b], all statistics shown in the report are based on EMIR trade-state data provided by all TRs and are presented as the number of contracts outstanding, or the notional value of contracts outstanding expressed in EUR after conversion. The data are as of 30 June 2023.

EMIR data are a vast source of detailed information on European derivatives markets. As these data span the entire European derivatives markets which are composed of a considerable number of market participants trading a wide range of asset classes and products, they are very voluminous and complex. This renders the necessary data cleaning and preparation procedures, to enable processing and aggregation, rather challenging. These procedures, such as outlier detection, are explained below and can be applied to other projects using the EMIR data set.

To ensure a high level of data quality, CSSF and CAA employed a multi-step data preparation procedure. The CSSF-CAA outlier removal approach relies on 3 steps. The first step applies two thresholds to the Luxembourg data: a fixed threshold and a dynamic one. The fixed threshold of notional amount of EUR 10bln results in the exclusion of reports which exceed this threshold, while the dynamic threshold results in the exclusion of reports whose log of notional amount exceeds the median plus four standard deviations of the distribution of the log of the notional amounts. As the market is very heterogenous the dynamic threshold is calculated for each cluster represented by the following fields: asset class, contract type, intragroup, compression and notional currency. The second step is on the application of European Union wide thresholds provided by ESMA [2023a] per asset classes, any notional amount above the median plus 4-times the standard deviation is disregarded from the analysis. The third and final step is the use of expert judgment which allows the identification and exclusion of some specific outliers per market segment.

Based on the supervisory analysis performed on the statistical data, CSSF and CAA acknowledged that the outliers are not only transactions of highest importance for supervisors but also transactions that have to be verified because they are either

- correct, and therefore to be included in the analyses; or
- wrong, and thus to be corrected by reporting counterparties (and implicitly removed).

2.4 Luxembourg market's key trends

According to the data provided by the TRs, some of the key trends from December 2021 to June 2023 are as follows.

- The Luxembourg derivatives market is relatively stable in terms of both structure and size, with a total notional of between EUR 6,000bln and EUR 7,000bln and between 850,000 and 950,000 transactions.

- In terms of notional amounts, currency derivatives and interest rate derivatives dominate the market, with 42% and 36% respectively of the total notional amount outstanding.
- In terms of number of outstanding derivative contracts, currency derivatives represent almost 50% followed by equity derivatives (34%) and interest rate derivatives which represent only 11% of transactions.
- Short-term maturities prevail in terms of notional with more than 70% of the derivatives having less than one-year remaining maturity.
- Not surprisingly, investment funds are the main participants in the derivatives market, accounting for more than 62% of trading activity in terms of both notional and number of outstanding transactions. Credit institutions represent slightly more than 10% in terms of notional and about 17% in terms of number of outstanding transactions.
- OTC derivatives significantly prevail representing more than 80% of trading activity in terms of notional compared to ETD.
- Interest rate derivatives are by far the most cleared asset class, representing nearly 70% of all cleared notional of OTC derivatives.
- EUR and USD are the most relevant currencies used to report notional amounts.
- The concentration of the Luxembourg derivatives market is relatively low with 905 counterparties representing 80% of the market in terms of total notional (see Figure 1).
- The interconnectedness of the Luxembourg derivatives market is quite stable, with approximately 4 connections per counterparty.
- The commodity asset class is the class where some deviation from the general trends have been observed. In particular, the remaining maturity below 1 year only decreased during 2022; it rose again almost to the previous levels in 2023.
- The interconnectedness level increased, with the average number of connections per counterparty rising from 5.5 in December 2021 to 8.5 in June 2023, while the trend for the others asset classes remained stable.

3 Assessing data quality: Monitoring and indicators

3.1 Data quality indicators

NCAs and ESMA have developed a common set of 19 Data Quality indicators (DQIs) to enable the detection and measurement of various types of misreporting.[5]

[5] These DQIs indicators were developed on the basis of pre-EMIR Refit data, similar indicators are currently under development using EMIR Refit data.

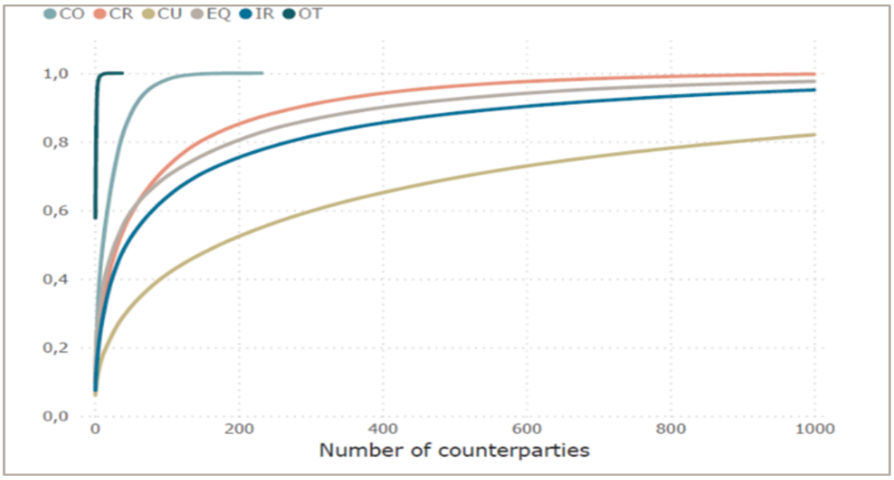


Figure 1 – MS11: Concentration – aggregated notional by asset class.

These indicators target different types of misreporting behaviours and are related to different types of misreporting.

Based on the specific indicator the behaviour underpinning the misreporting is categorised on 3 dimensions (Figure 2).

- **Misreporting by reporting entity.** These indicators allow one to identify and measure a clear misreporting by the entity responsible for the report.
- **Misreporting by either one of the entities.** These indicators are related to the double-sided reporting obligation introduced by EMIR where both counterparties to a derivative contract are required to report the contract details and therefore shall agree on the details to be reported in order to ensure a perfect match. As per experience gathered over the last decade, two sub-categories of indicators have been created:
  - pairing and matching as performed by Trade Repositories;
  - comparison of the information reported by both counterparties.
- **Potential misreporting.** These DQIs identify behaviours which are highly probable to be representative of misreporting situations. However, it could be possible that the indicator captures wrong positives, i.e. where the reported data appear to be misreported but in fact are correctly reported as per the accurate contract details.

3.2 Data quality monitoring

DQIs amplify their supervisory power when looked at in a comprehensive way either at national level, or at sectoral level, but mainly at entity level.

The first pillar of the new strategic approach is a comprehensive data quality dashboard (Figure 3) to allow for a consistent monitoring of the evolution of the quality of a given dataset over time. The data quality indicators, when applied to

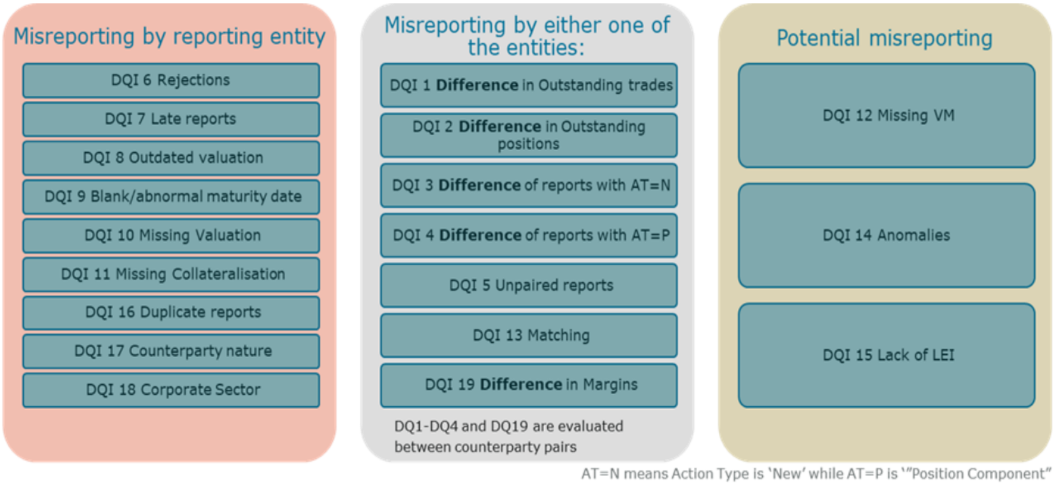


Figure 2 – DQI list by typology of misreporting, CSSF.

a given country or reporting entity, facilitate a comparison of the quality of that reporting against that of the EU market. Furthermore, tracking the entities' results over time can also be used to measure, in an objective manner, the effectiveness of the supervisory activities undertaken and the improved level of data quality.

The EMIR data quality dashboard was agreed in May 2022 and gradually has been implemented since then. The 19 DQIs are used to detect and measure various types of misreporting, including under and over reporting, inconsistent reporting between the two counterparties to the transaction (where both are required to provide data to TRs), incomplete information in the key fields of the reports, late reporting, abnormal values, and lack of correct identifiers of the counterparties. The DQIs are computed on a monthly basis based on the EMIR dataset. Significant reporting irregularities are followed up in a systematic manner under the agreed NCA engagement framework.

3.3 NCAs data sharing frameworks

The second pillar of the approach is a common framework for the provision of data and follow-up on significant data quality issues. The main goal of this framework is to ensure that the resolution of the most critical data quality problems is performed as swiftly as possible and with an efficient use of NCAs' and ESMA's resources.

In particular, the framework specifies the criteria which should be used to determine which reporting issues should be considered significant and prioritised as well as which entities should be targeted in the follow-up based on the quality of their reporting. The important feature of the framework is that the follow-up is focused on a limited subset of entities with the highest volume of incorrect reports at EU level, thus ensuring the most efficient use of the NCAs resources. Under certain circumstances individual entities may be approached, e.g. when they report abnormal/incorrect values on such a scale that it may materially impact the analysis of EMIR data. Thus, the framework follows a risk-based approach.

Figure 3 – Data Quality Comparison between Luxembourg and European Markets.

3.4 Considerations and lessons learned regarding data quality

The inadequacy of data quality is a significant hurdle for the extensive use of data reported by entities to extract meaningful information on derivatives markets. While data cleaning procedures are implemented to address these shortcomings, they often carry the risk of overlooking genuine anomalies; despite their intended purpose of enhancing data quality.

In addition, access to data and indicators should be given not only to those dealing with EMIR data, but also to those in charge of the supervision of entities. To facilitate this broader use of EMIR data among entity supervisors, it is essential that the information is presented in a user-friendly format. All stakeholders, particularly those counterparties involved in derivatives trading activity, are encouraged to use and leverage on the regulatory data in their internal processes. This collaborative approach can foster a symbiotic relationship between improved data quality and aligned incentives. It holds the potential to cultivate a more robust and trustworthy trading environment.

For supervisors, it is important to have a clear indication of the overall data quality of an entity, a sector, and the market as a whole to deliver the appropriate messages to entities, industry and market associations. The ultimate goal of this collective endeavour is to bolster data use across the board, empowering all data users to effectively fulfill their respective responsibilities.

Bibliography

CSSF. EMIR Refit reporting standards. https://www.cssf.lu/wp-content/uploads/PR22_33_EMIR_Refit_reporting_standards_211222.pdf, 2022.

CSSF and CAA. CSSF/CAA joint conference on EMIR reporting. https://www.cssf.lu/en/2023/06/cssf-caa-joint-conference-on-emir-reporting/, 2023a.

CSSF and CAA. Luxembourg derivative market 2023: Joint CSSF-CAA report. https://www.cssf.lu/en/Document/luxembourg-derivative-market-2023/, 2023b.

ESMA. ESMA 2023 report on quality and use of transaction data. https://www.esma.europa.eu/sites/default/files/2024-04/ESMA12-1209242288-852_2023_Report_on_Quality_and_Use_of_Data.pdf, 2023a.

ESMA. ESMA derivatives markets 2023. https://www.esma.europa.eu/sites/default/files/2023-12/ESMA50-524821-2930_EU_Derivatives_Markets_2023.pdf, 2023b.

ESRB. 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, 2022a.

ESRB. Letter on ESRB view regarding data quality issues and risks for financial stability, published on 13 July 2022. https://www.esrb.europa.eu/mppa/responses/html/index.en.html, 2022b.