

EMIR data for financial stability analysis and research

Dario Ruzzi

(joint work with M.L. Bianchi and B. Sorvillo)

Banca d'Italia

Financial Stability Directorate

3rd IFC and Bank of Italy Workshop on

Data Science in Central Banking: Enhancing the access to and sharing of data

October 18, 2023

Outline

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

EMIR data in a snapshot



EMIR data

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

The information collected in EMIR can be categorized as follows:

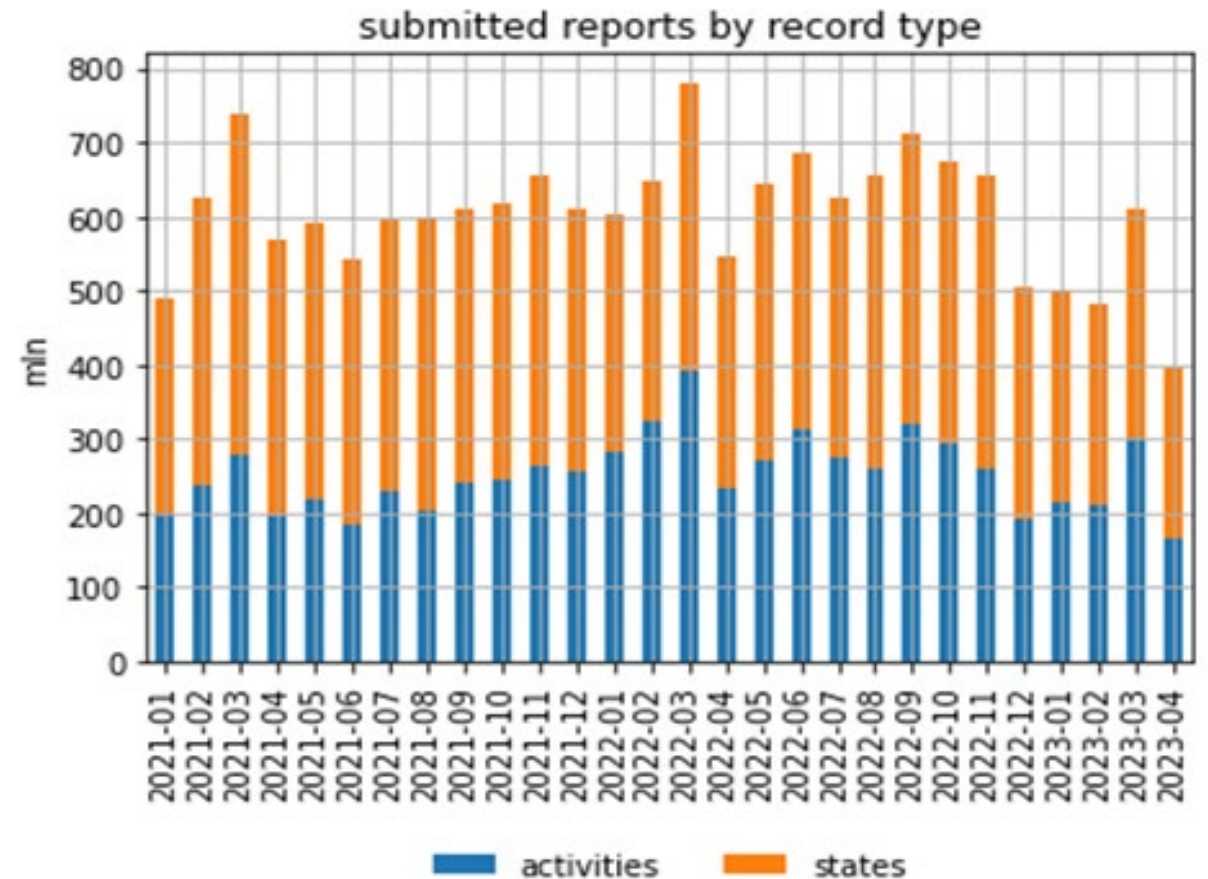
- *trade details*: detailed information related to derivative transactions, including trade identifiers, dates and times of trade execution, product type, underlying assets, notional amounts, contract maturity dates, contract characteristics;
- *counterparty information*: counterparties involved in derivative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

EMIR data

The reports are of two types:

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

For risk analyses purposes “states” are usually considered.



Access to data

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



Use of data

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

Some data quality issues

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

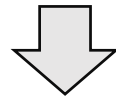
Data cleaning: our steps

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

Deduplication: our approach

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

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

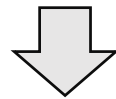


cleaned deduplicated data
ready for the analysis

Dealing with collateral portfolios

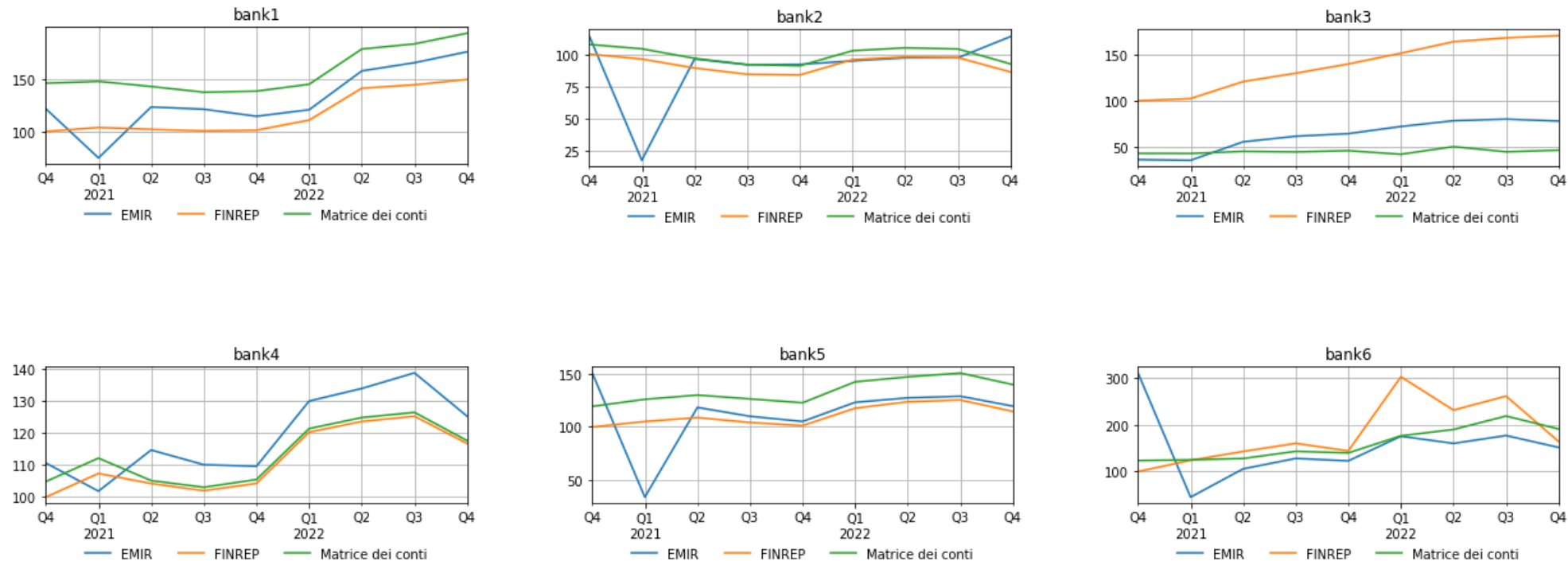
Margins are reported at the collateral portfolio level:

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



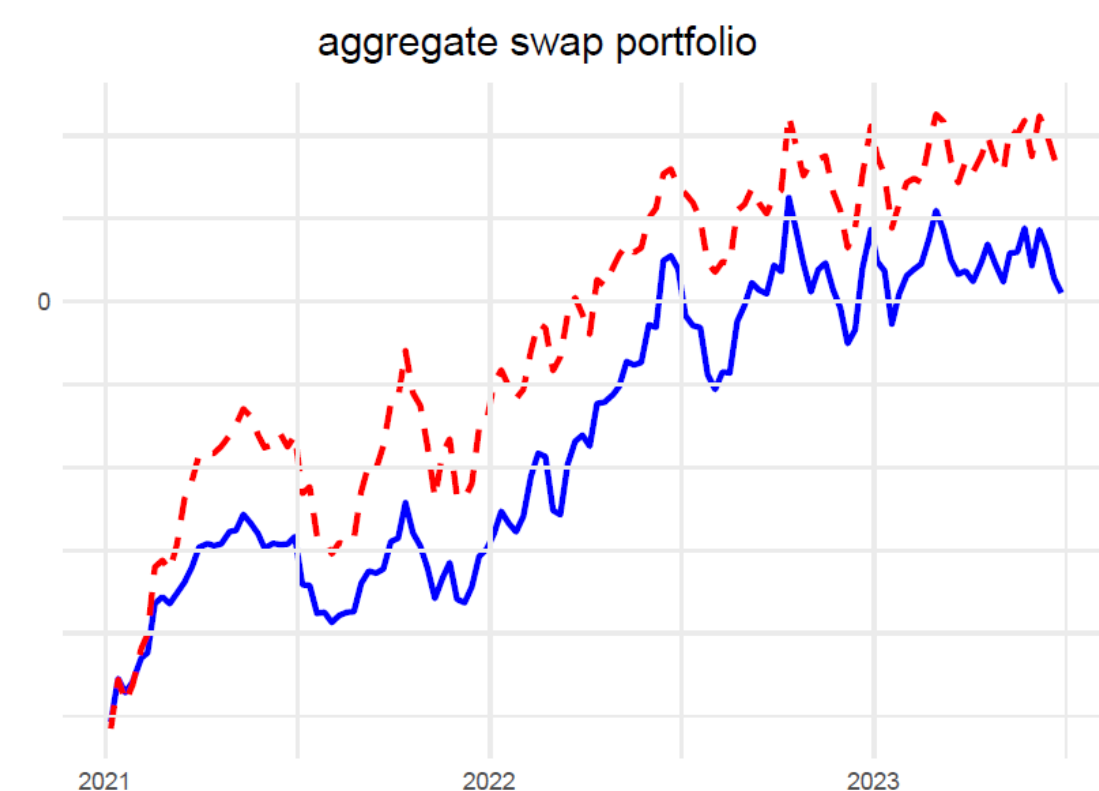
we NEED the double reporting here!

Data quality on notional values: EMIR vs Finrep

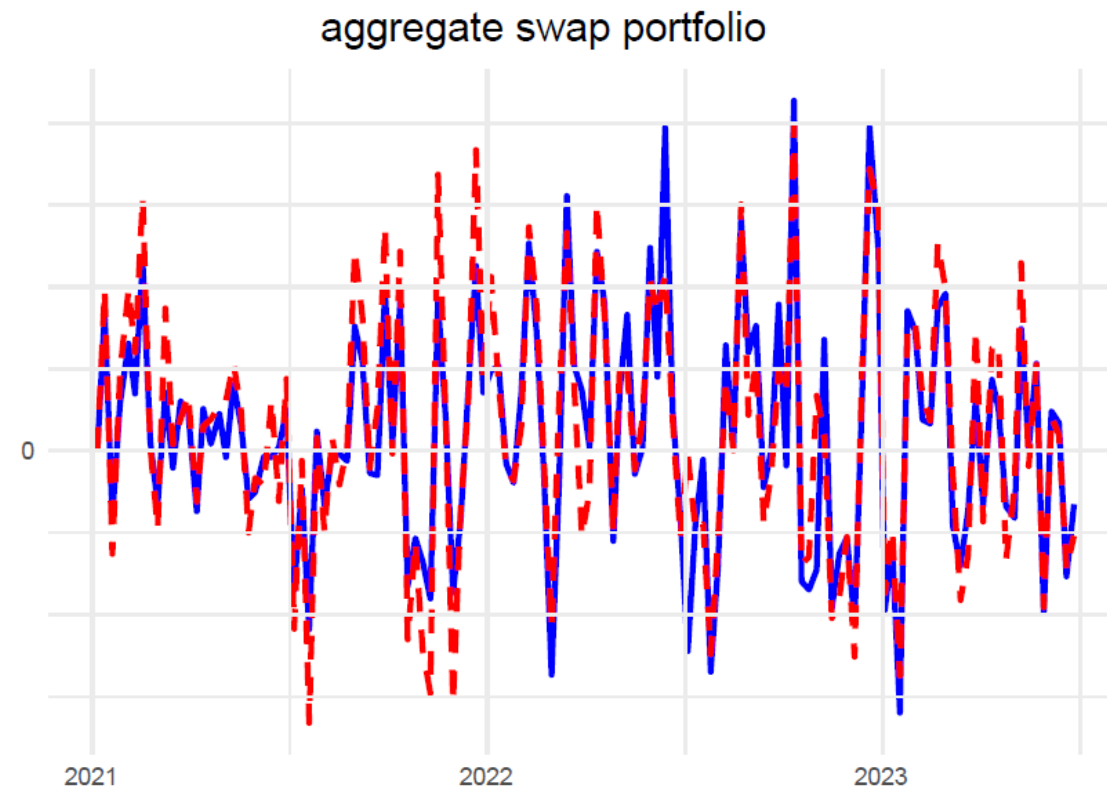


- Source: cleaned deduplicated EMIR data available at the Banca d'Italia, FINREP and supervisory and statistical reports (Matrice dei conti). All data are rescaled so that, for each bank, the FINREP value is equal to 100 at the Q4 2020.

Data quality on the value of contract: IR swaps



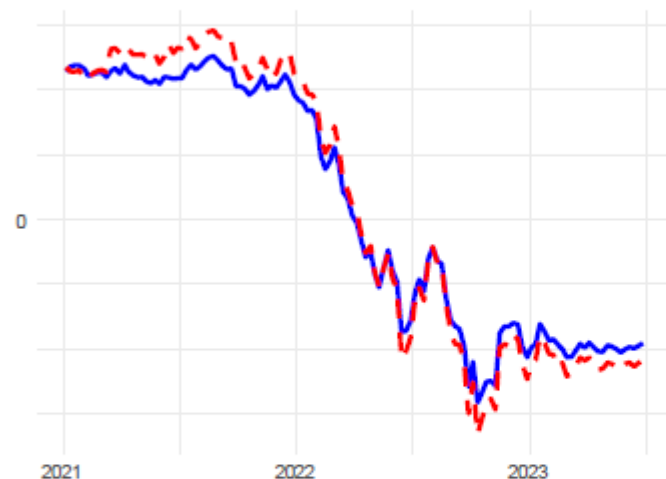
swap portfolio observed in EMIR and implied by the model



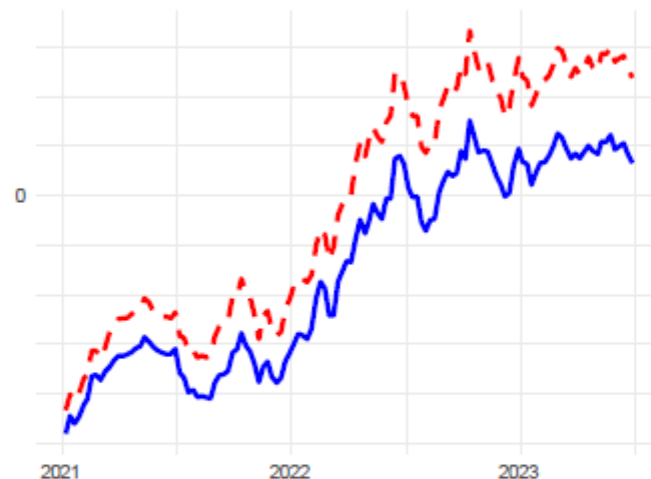
swap portfolio observed in EMIR and implied by the model

Data quality on the value of contract: IR swaps

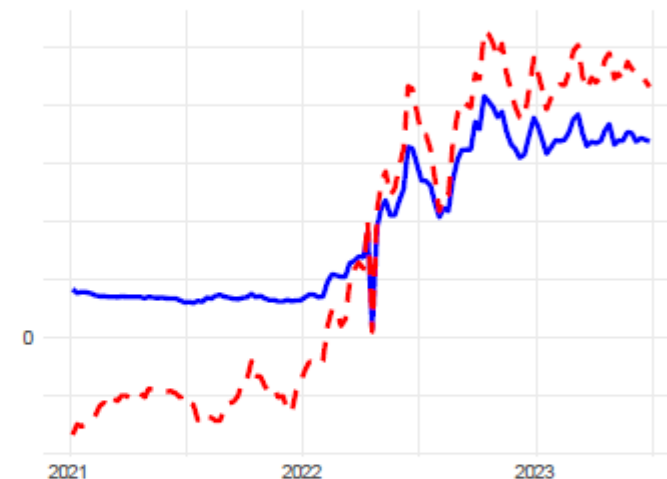
the Good



the Bad



the Ugly



swap portfolio **observed in EMIR** and **implied by the model**

Research works on EMIR data at Banca d'Italia

- *Shifting the yield curve for Italian banks*: valuation of interest rate derivative contracts and risk sensitivity analysis using granular information on these derivatives and fixed-rate bonds (*available upon request*)
- *CO₂ emission derivatives In Italy*: analysis of the market of CO₂ emission permits derivatives in Italy (*available upon request*)
- *EMIR networks*: investigation of the structure of EMIR networks (notional and margins), their dynamics over time and the main risk drivers (*work in progress*)
- *Interest rate futures*: analysis on sovereign bond futures (*work in progress*)
- ... a lot of work

Thank you for your attention