EUROPEAN CENTRAL BANK

EUROSYSTEM

Grzegorz Skrzypczynski

Sébastien Pérez-Duarte

DG-Statistics, European Central Bank

Two is company, three's a crowd:

Automated pairing and matching of two-sided reporting in EMIR derivatives' data

IFC conference

Are post-crisis statistical initiatives completed?

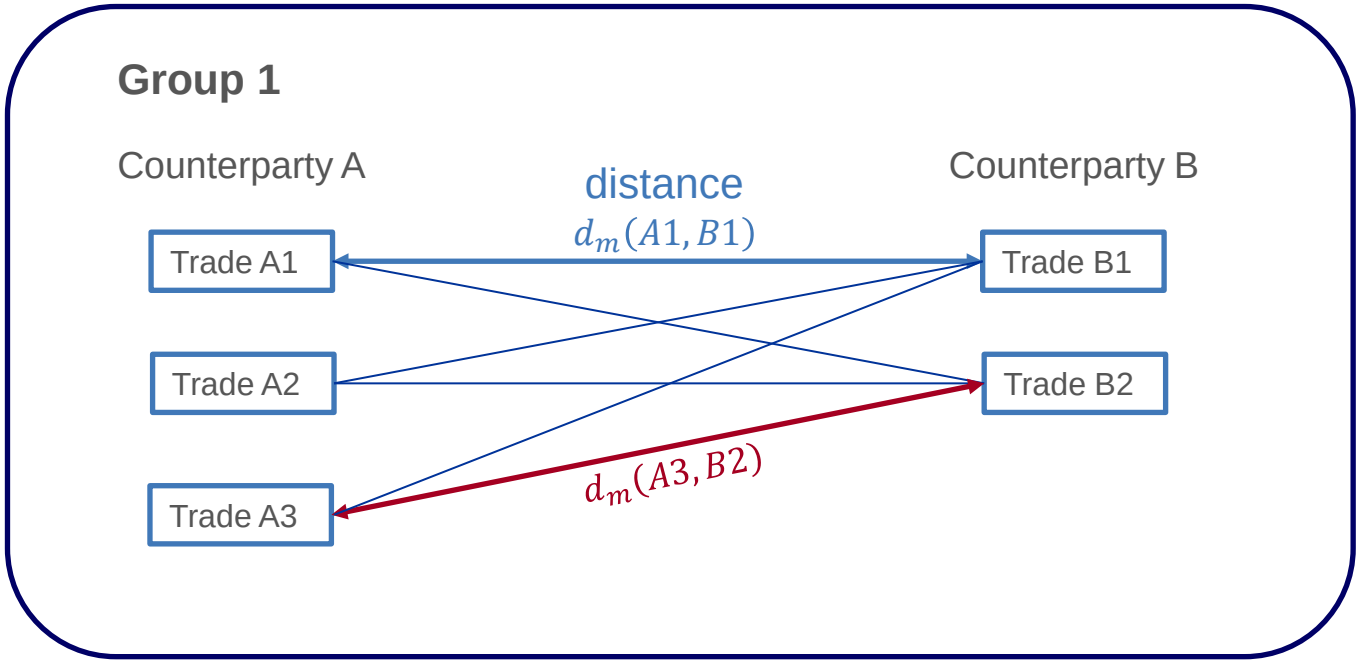
30-31 August 2018, Basel

Motivation: double-sided reporting in EMIR

- EU counterparties report **derivative transactions** to trade repositories
- Separately by both counterparties (**double-sided reporting**)
  - Improves data quality monitoring
  - ⓘ Risk of double-counting when analysing and aggregating the data
- **UTI (Unique Transaction Identifier)** to link trades, agreed between counterparties
  - Challenges in implementation (not unique, different UTIs for the same trade)
  - Work on improving pairing and matching (inter-TR reconciliation process)
  - Global initiatives to harmonise UTI between jurisdictions
  - ESMA estimates pairing rate at 87% = but newly reported trades only

Method

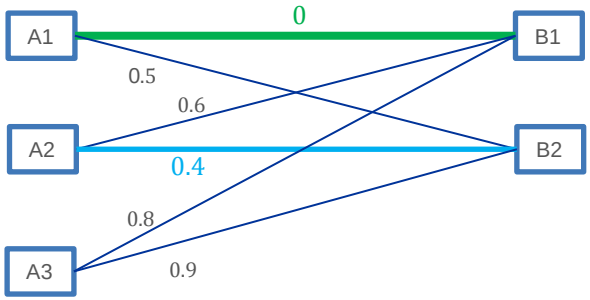
- The trades are split into groups with same values of the **grouping** variables
- The procedure calculates the **matching distance** between each member trade of the group



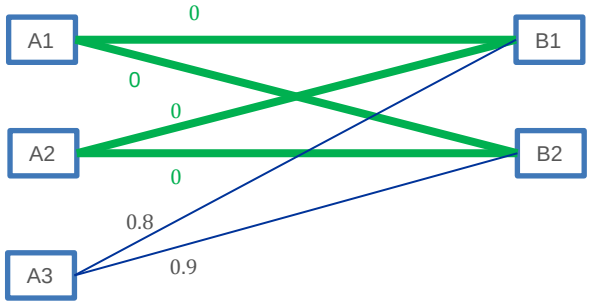
Classification of trades

- Depending on outcome, exclusive categories

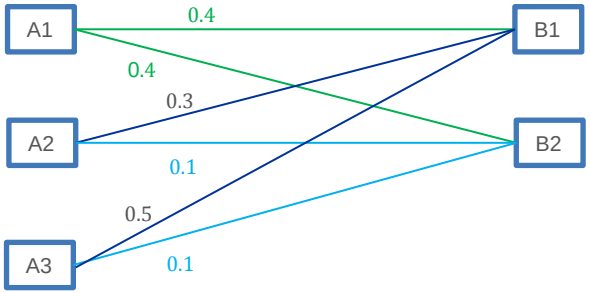
Perfect match / imperfect match



Perfect matching group



Ambiguous



No match

Implementation

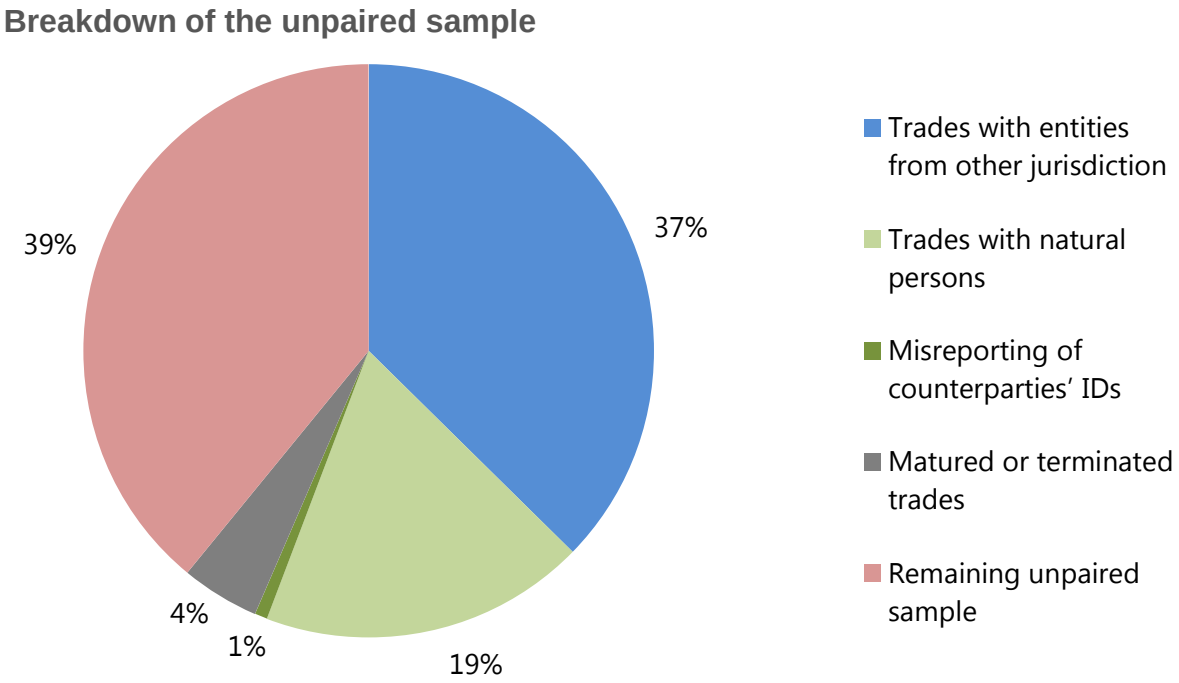
- Sample paired with UTI was used to calibrate the parameters
- **Grouping variables**:
  - Counterparties' IDs
  - Asset class
  - Contract type
  - Clearing status
  - Execution date
- **Matching distance weights**: function of fidelity (how good) and specificity (how revealing) of the variable
- **Thresholds of the distance function**: to accept 98% of the observations in the paired sample

Implementation

- Our procedure has limited impact in the non-paired sample
- Most of these trades don't have any counterpart in their group
  ➔ other reporting issues are at stake

	Paired sub-sample	Non-paired sub-sample
Perfect match	48.74%	0.49%
Imperfect match	28.39%	2.73%
Perfect matching group	12.99%	0.08%
Ambiguous	5.53%	3.61%
No match	4.35%	93.09%

Some things we will never be able to pair



- Remaining 40% of the unpaired sample:
  - may be a result of underreporting
  - may constitute invalid reports

What can't we pair?

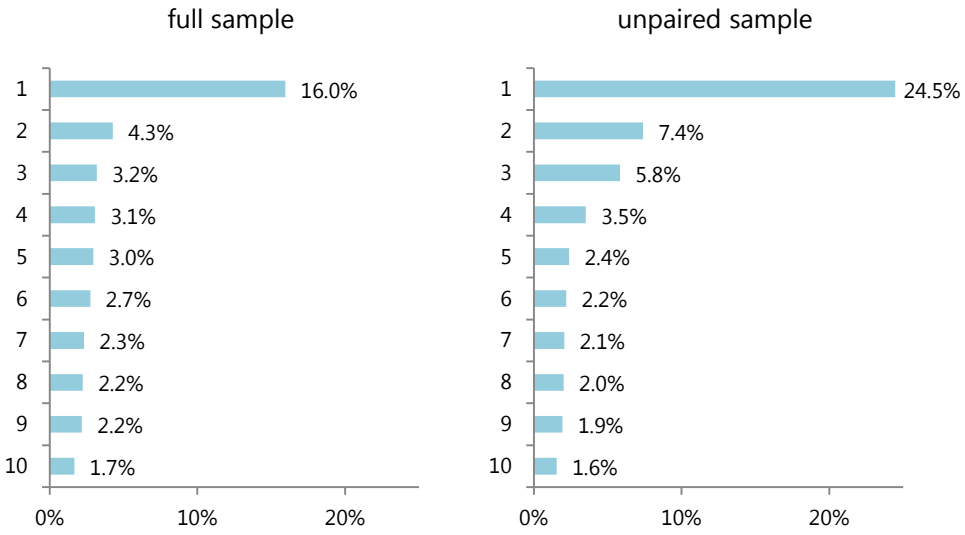
Logit regression - Odds ratios, probability of being paired

Variable	Odds ratio
Asset class	
Commodity	1
Credit	0.547
Currency	0.508
Equity	0.619
Interest rate	0.745
Other	0.324
Missing	1.371
Contract type	
Contracts for difference	1
Forwards	1.906
Forward rate agreements	1.919
Futures	0.879
Option	1.197
Other	0.420
Swap	1.770
Swaption	1.899
Missing	0.173
Clearing Status	
No	1
Yes	1.720
Missing	0.439
Intra group	
No	1
Yes	4.228
Missing	0.669

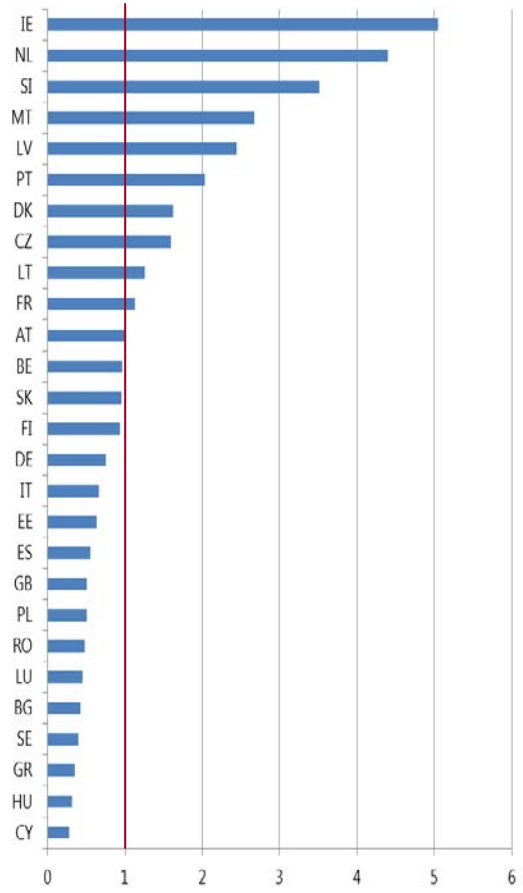
Variable	Odds ratio
Location of other counterparty	
Euro area	1
Other EU	0.405
RoW	0.00676
Nature of reporting party	
CCP	1
Financial	1.378
Non-financial	4.740
Other	2.600
Missing	5.220
Execution date	
<= 2013	0.247
2014-2017	0.403
2017	0.645
2018 Q1	1.178
2018 Apr-May	1.197
2018 Jun	1
2018 Jul	0.599
> Aug 2018	0.608
Contract value missing	
No	1
Yes	0.590
Notional amount (log)	1.036

What can't we pair?

Share of largest reporters in the unpaired and the full sample



Pairing success by country of the other counterparty *(odds ratio)*



Conclusions

- Caution is recommended when **making assumptions** about the unpaired sample **to compute aggregates**
- A **significant share of the non-paired sample is difficult to interpret,** and cannot be easily reconciled
- There exist some **clear patterns** between some characteristics of the contracts and **probability of being paired**
- The unpaired sample exhibits **higher concentration** with regards to reporting entities
- A **focused data quality management** process may bring **significant benefits with limited effort**