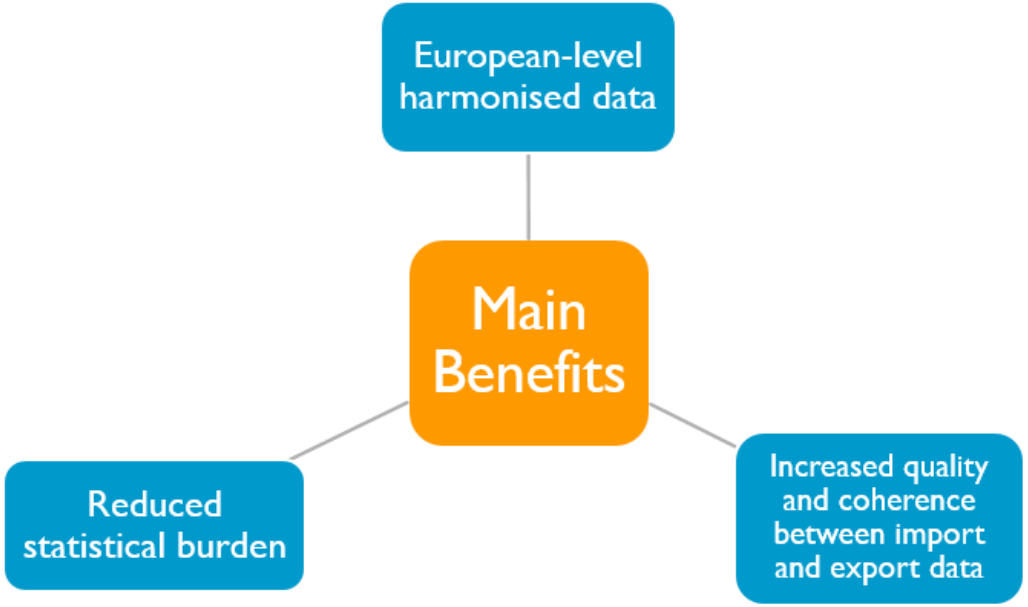
R Shiny Apps for asymmetry analysis via selective editing

M. Bruno, M. S. Causo, G. Massacci, F. Ortame, G. Ruocco, S. Toti

Context

Intra-EU microdata exchange (MDE) provides statistical institutes with a new data source for compiling and benchmarking intra-EU imports



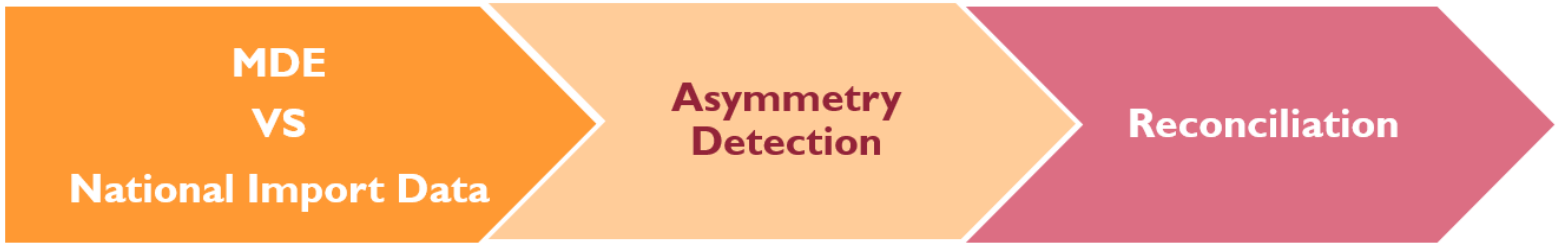
Microdata exchange is based on the principle that data relative to the same phenomenon should be collected only once.

Goals

Standardized approach to data editing in the context of the European Statistical System (ESS):

- Implement a standardized and harmonized statistical production system that exploits MDE data to detect and reconcile **asymmetries** between import and export data
- Analyze the discrepancies between the MDE data source and national data sources

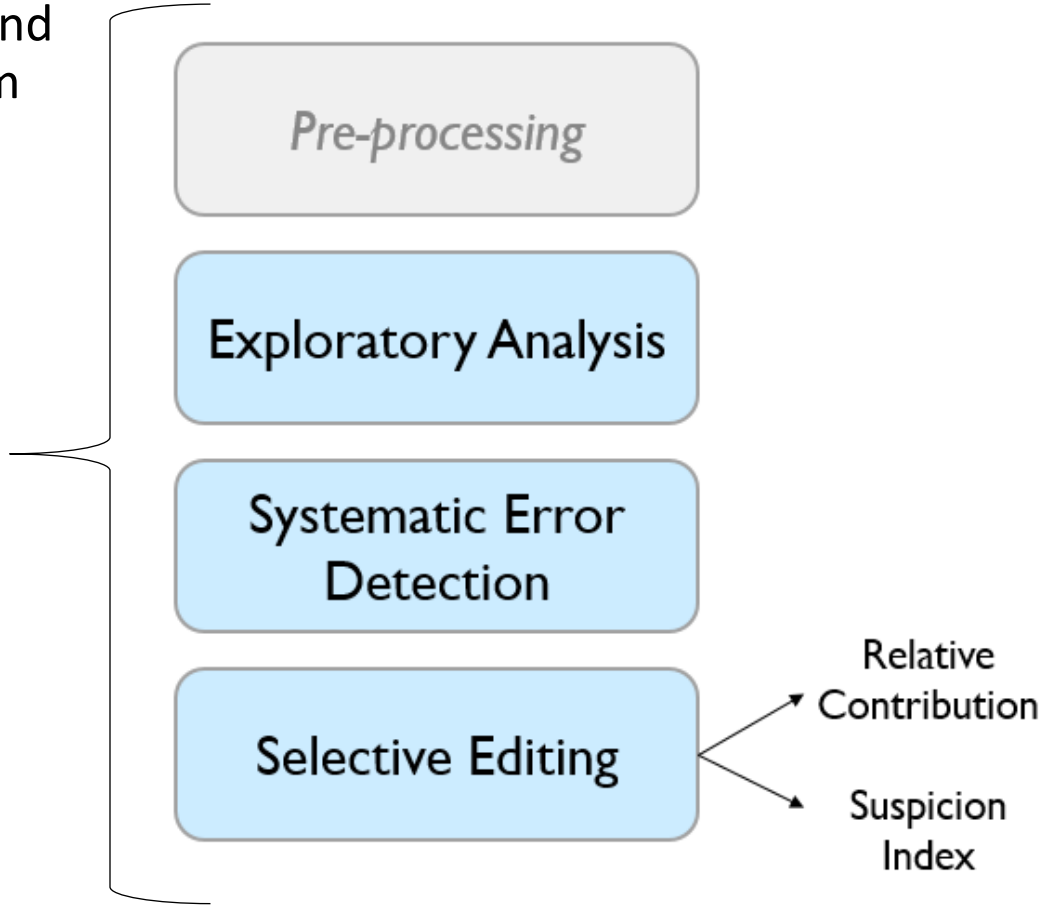
Develop a European-level harmonized tool to tackle intra-EU statistical asymmetries



Main principles

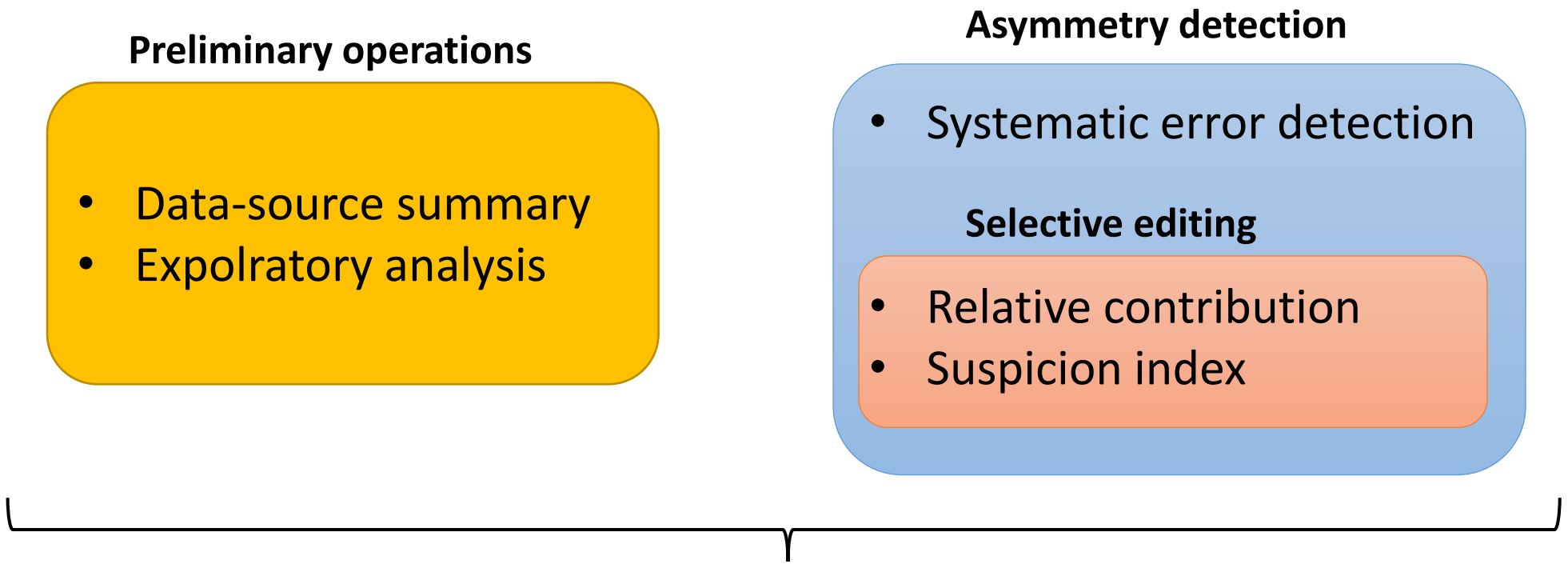
- Implement shareable methods for asymmetry analysis through an **open-source** solution
- Work towards Highlighting the **most influential units** and systematic errors through a combined threshold system
- Facilitate reconciliation for the detected asymmetries
- Build a **taxonomy** for asymmetries to identify the most appropriate validation methods

Following the Generic Statistical Data Editing Model (GSDEM) guidelines, we developed a tool that includes:

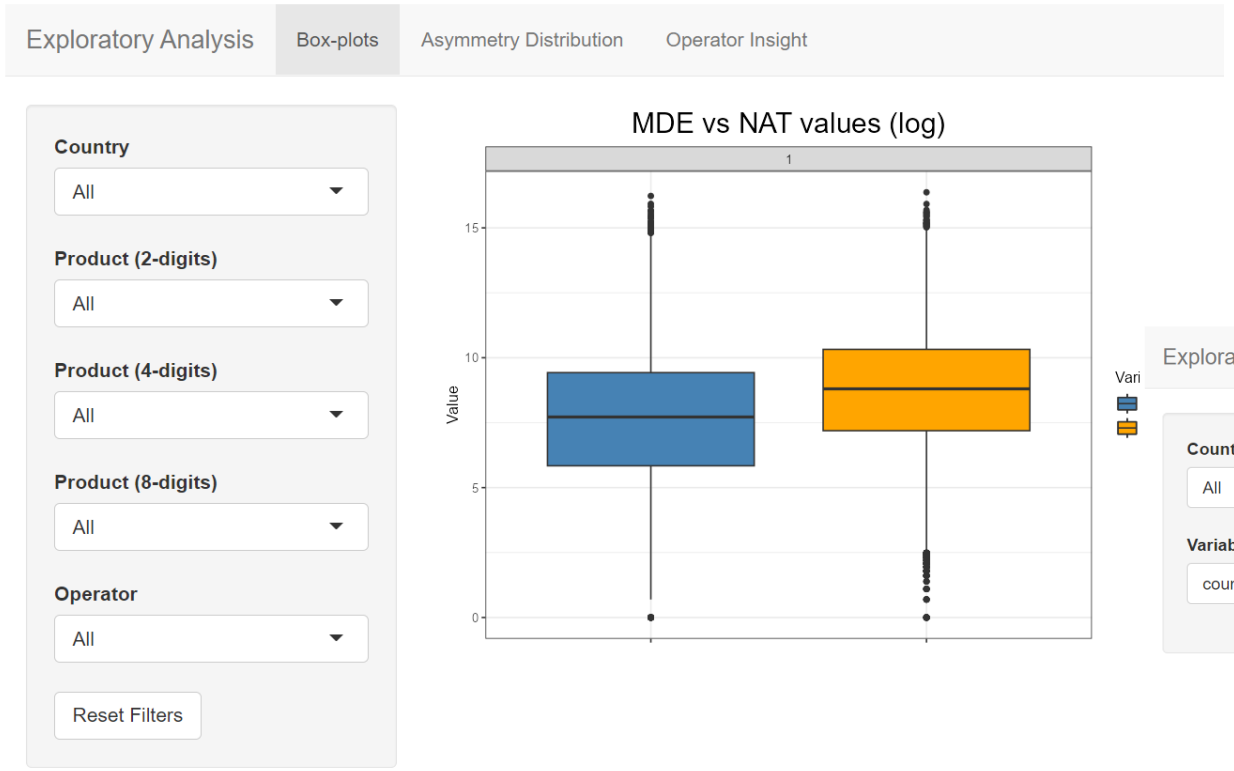


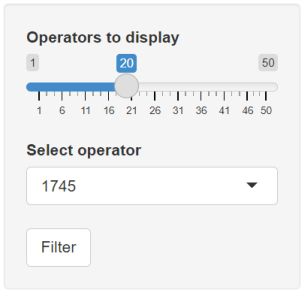
AsyD – Asymmetry Detection

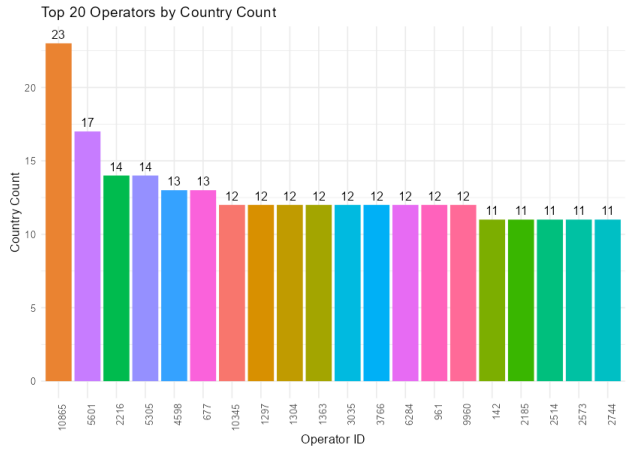
Open-source tool developed to identify the most relevant asymmetries

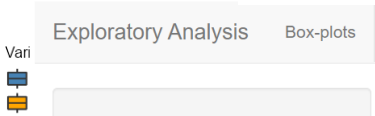


Exploratory Analysis









Asymmetry Detection (Systematic Errors)

CSV

Search:

	operator_id	value	product_id_8_nat	country_id_nat	product_id_8_mde	country_id_mde
583	5609	504324	1022999	ES	1022951	ES
383	4023	494945	1039190	DE	1039219	DE
200	226	417853	4069073	DE	4069089	DE
214	2357	333838	4069001	LV	4061080	LV
703	6866	306380	1022991	FR	1022949	FR
513	5114	298694	4061050	DE	4069089	DE
172	2066	280512	3074338	ES	3074335	ES
350	367	254014	5051090	DE	5051010	DE
652	6294	245071	1022959	SI	1022999	SI
584	5685	244306	4015039	AT	4041054	AT

Showing 1 to 10 of 1,036 entries

Previous 1 2 3 4 5 … 104 Next

Asystematic error is identified when the same value for MDE and national data sources is associated to a different product code

Asymmetry Detection (Selective Editing 1)

Relative contribution

The relative contribution to the total asymmetry is defined as follows:

$$C_i = \frac{(val_{MDE,i} - val_{NAT,i})}{\sum(|(val_{MDE,i} - val_{NAT,i})|)} * 100$$

The index is sensitive to real time groupings applied to the data frame

Suspicion index

The suspicion index is defined as follows:

$$S_i = \begin{cases} \frac{Q_1 - \log R_i}{Q_3 - Q_1}, \text{ if } \log R_i < Q_1 \\ \frac{\log R_i - Q_3}{Q_3 - Q_1}, \text{ if } \log R_i > Q_3 \\ 0, \text{ otherwise} \end{cases}$$

Where Q1 and Q3 are the 25th and 75th percentiles of the distribution of the difference between MDE and national values in log scale

Asymmetry Detection (Selective Editing 2)

Group by

country_id

Threshold for |contr|:

0,8

Mean: 8.58
Median: 3.41
SDev: 10.09

Filter

Relative Contribution Table

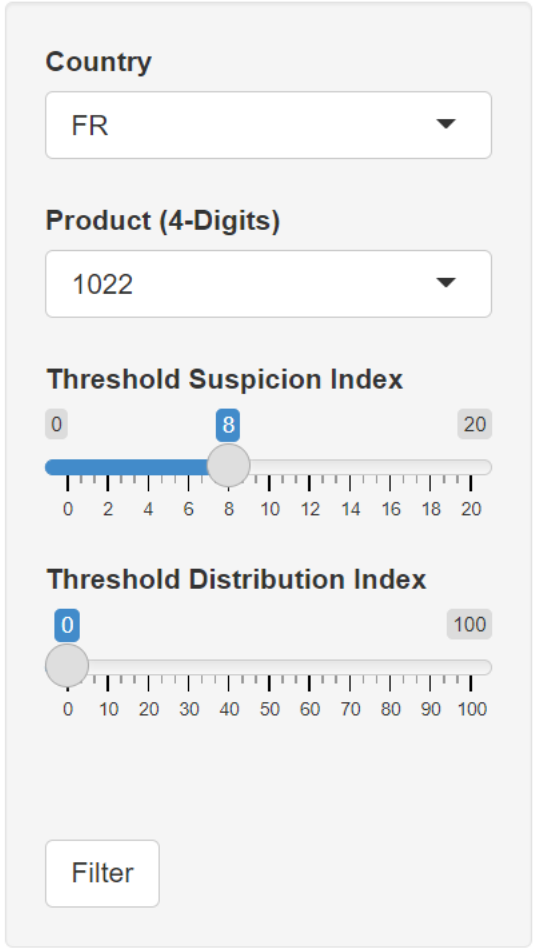
Column visibility ▾ | CSV | **Search:**

country_id	mde_value	nat_value	diff	contr
All	All	All	All	Al
AT	40890872	45189730	-4298858	-1.84
BE	62587890	74738475	-12150585	-5.201
DE	277973062	302256623	-24283561	-10.394
DK	62178202	70144546	-7966344	-3.41
ES	300268676	259803872	40464804	17.32
FR	311636517	270624445	41012072	17.554
HR	20471578	18200249	2271329	0.972
IE	27578756	34704183	-7125427	-3.05
NL	139791697	215766241	-75974544	-32.519
PT	10979791	8760295	2219496	0.95

Showing 1 to 10 of 11 entries

Previous 1 2 Next

Asymmetry Detection (Selective Editing 3)



Suspicion Index Table

Column visibility | CSV | Search:

operator_id	suspicion_index	distribution_index
All	All	All
1496	12.93	32.30
2905	8.40	59.91
7310	13.05	46.42

Showing 1 to 3 of 3 entries

Previous 1 Next

Final Remarks

- Limits of the current implementation:
  - Issues with server-side computing
  - Longitudinal analysis (asymmetry persistance)

- Cooperation with other member states to reach a harmonized solution

GitHub Repo:
https://github.com/istat-methodology/asyd