EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS

2024 ESTORIL - PORTUGAL

Instituto Nacional de Estatística
Statistics Portugal

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Measuring the quality of administrative sources: at macro level with novel indicators and micro level with distributions comparison

Alicia Nieto, Sandra Barragán, Alba Rodríguez,
Soledad Saldaña, David Salgado

S.G. for Methodology and Sampling Design
Statistics Spain

June 2024

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

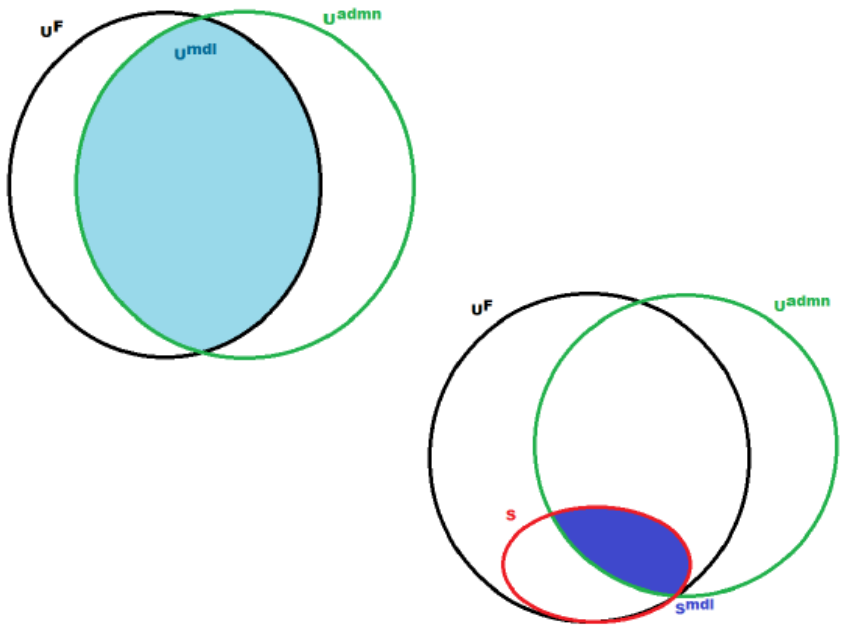
Introduction, background and purpose: admin-based SSAI

We are presenting our efforts to **replace** the traditional fully survey-based **statistics** Service Sector Activity Indicators (SSAI), with a combination of **survey and administrative data** from the Spanish National Tax Agency.

Our access to survey and administrative data from 2019 to the present has allowed us to conduct comparisons not only at the aggregate level but also at the **microdata** level.

The conference is partly financed by the European Union

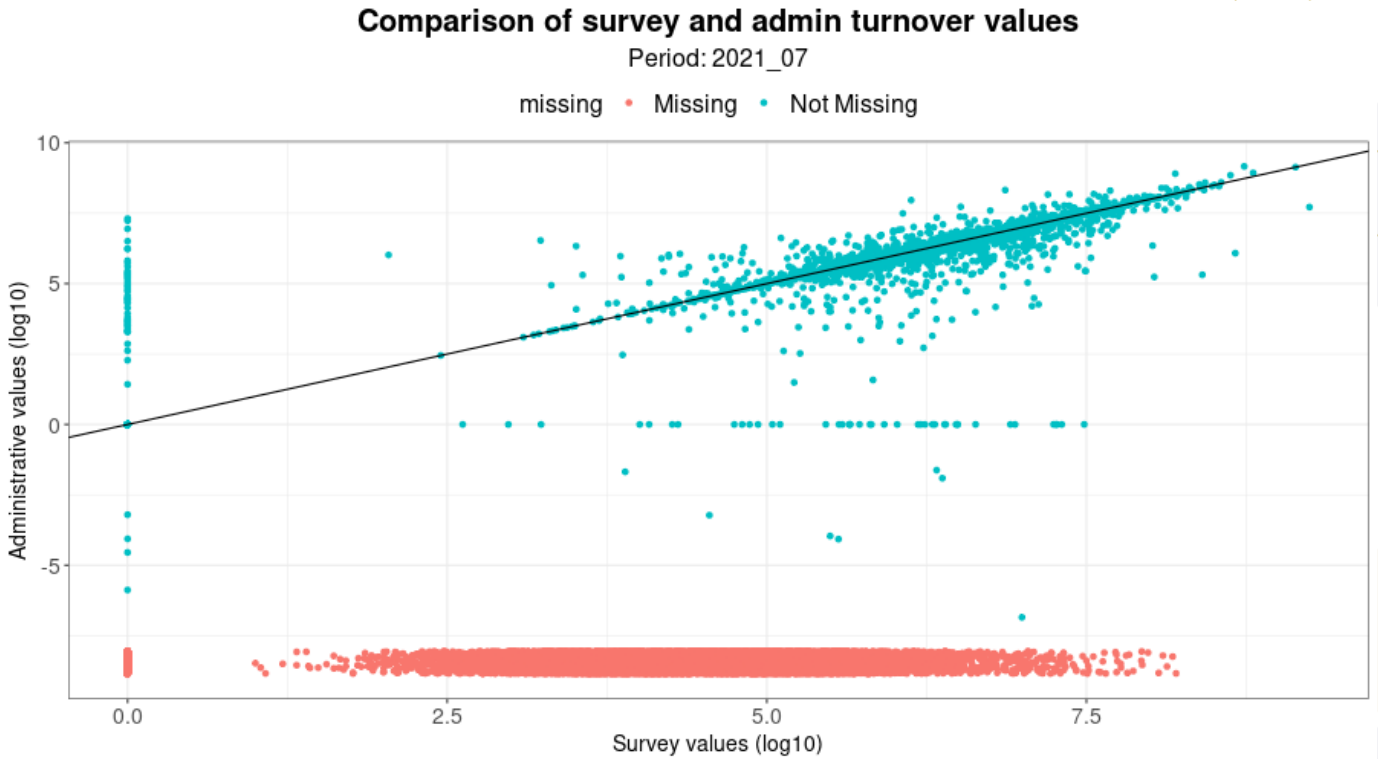
Data description



- U the **finite population** of analysis.
- $\mathrm{U_F} \subset SBR$ the **frame population** from the Statistical Business Register (SBR).
- U^{adm} the set of business units contained in the **tax register**.
- $U^{mdl} = U^{adm} \cap U_F$ the set of **statistical units** in the **tax** register.
- $S^{mdl} = S \cap U^{mdl}$ the set os statistical units in the training data set.
- y is the statistical variable of interest (**turnover** in this use case).
- y^{stat} denotes the values obtained with the **survey**.
- y^{adm} denotes the values obtained with the **administrative** register.

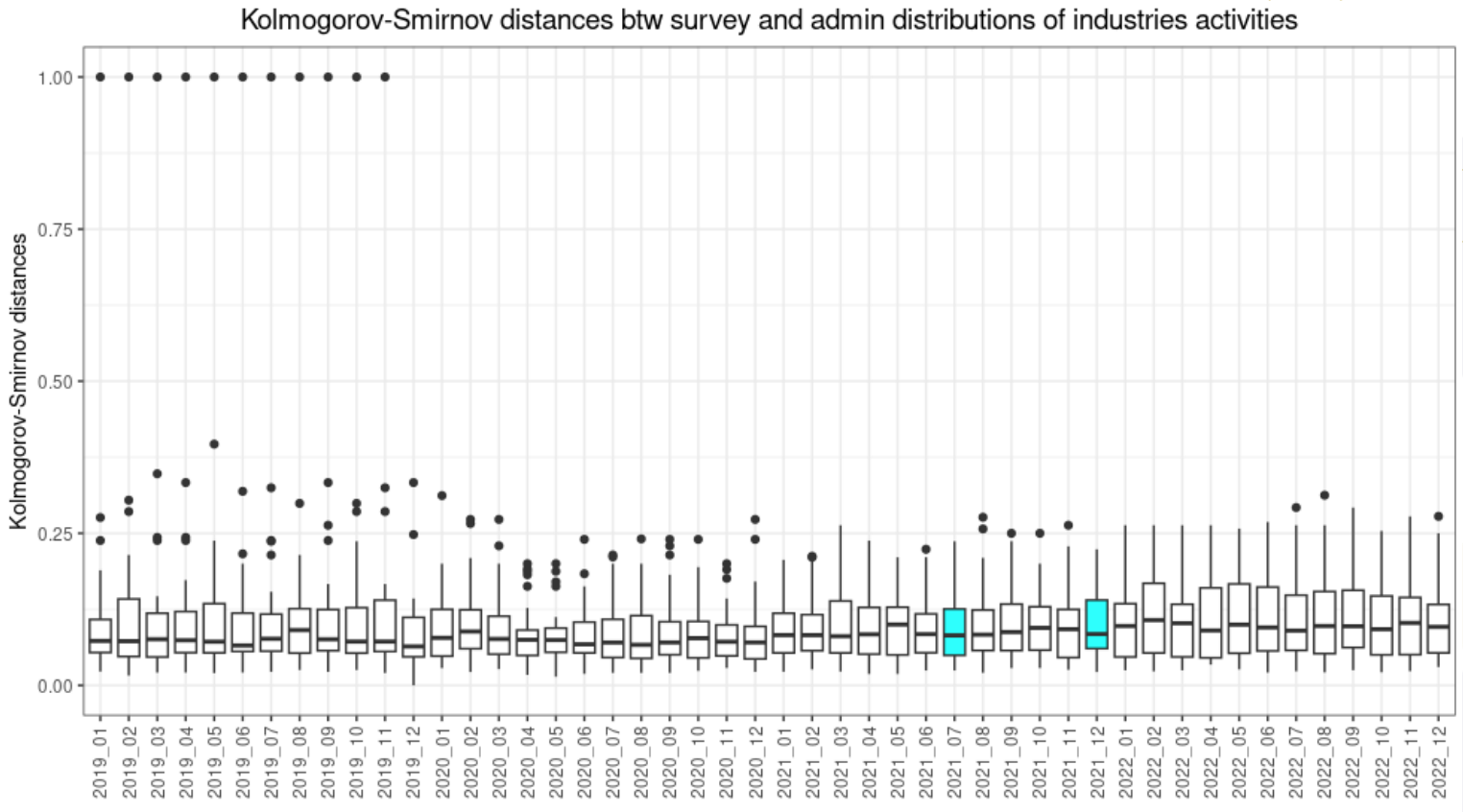
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Data comparison for unit-level microdata: scatterplot



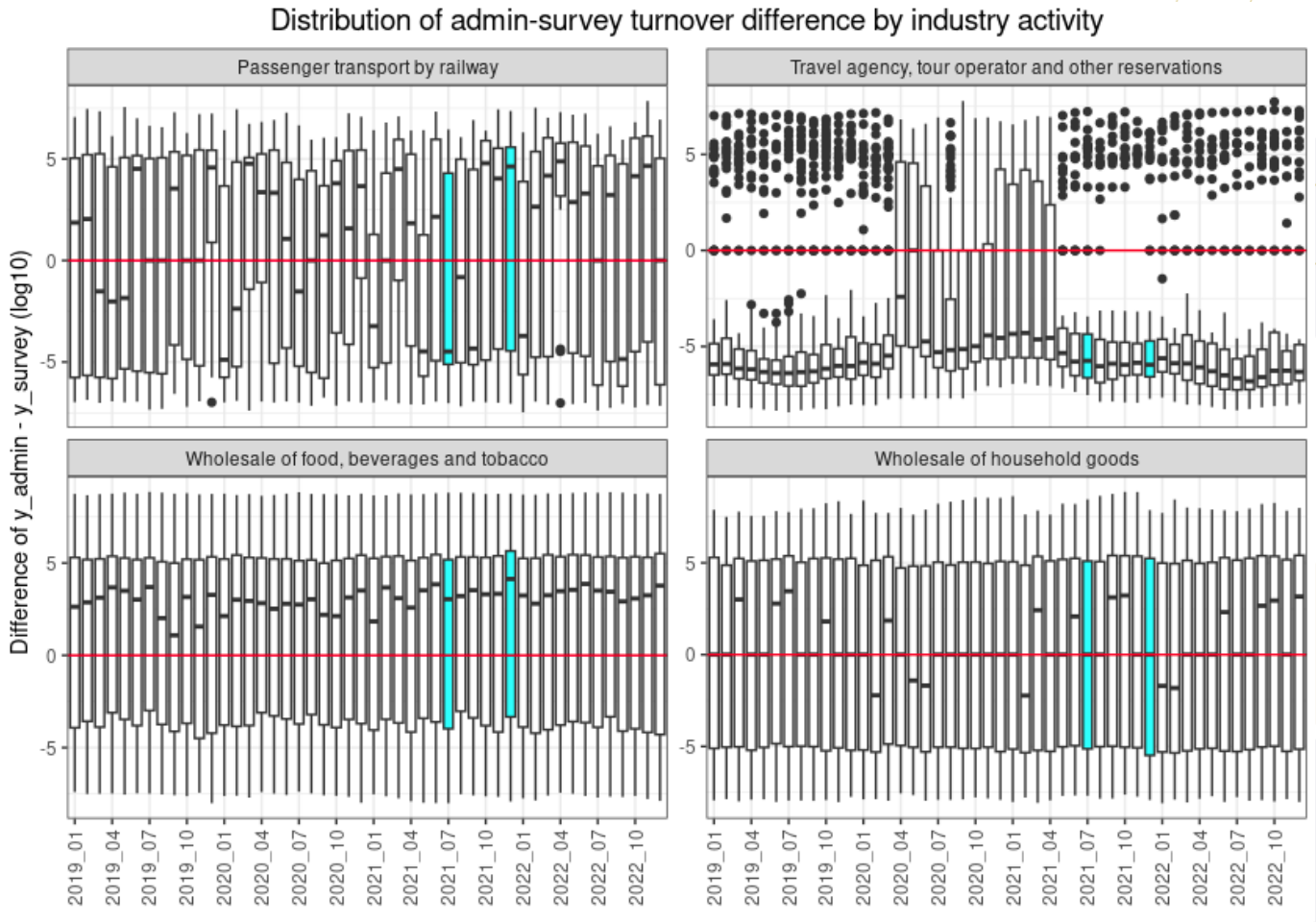
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Data comparison for unit-level microdata: Kolmogorov – Smirnov distance



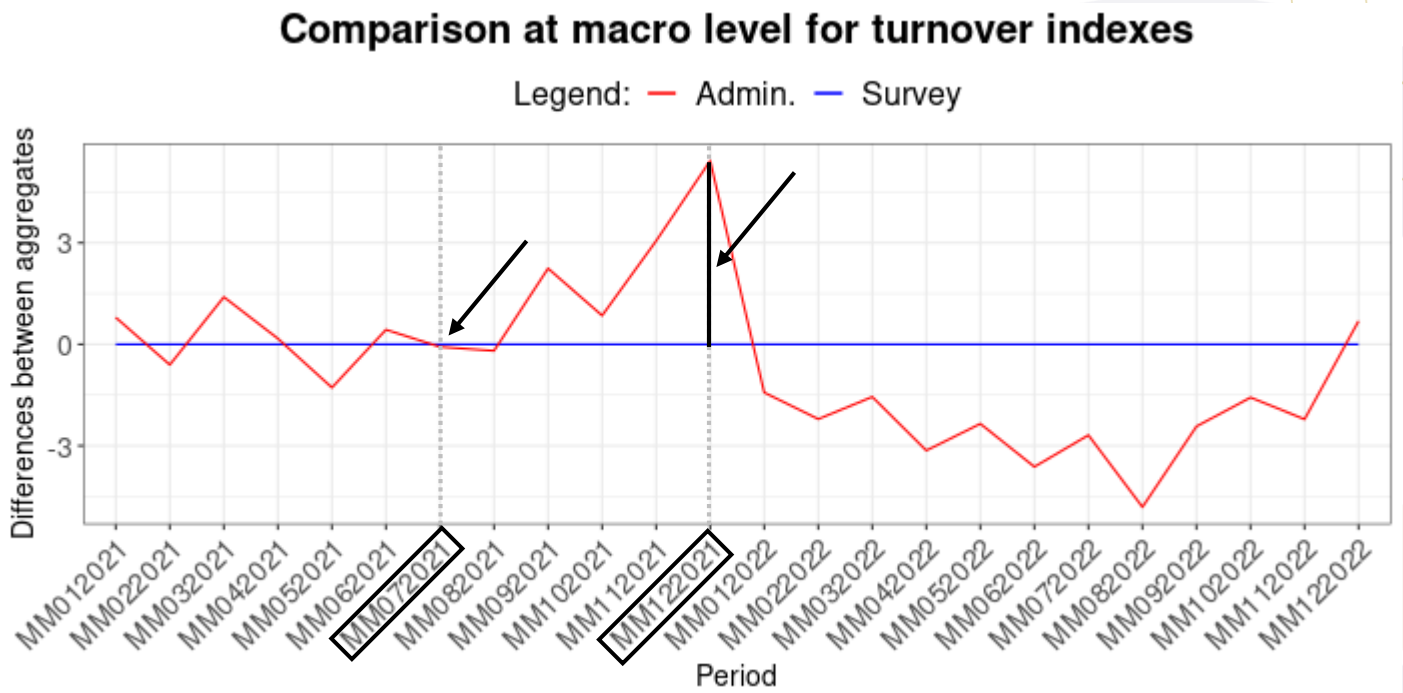
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Data comparison for unit-level microdata: turnover differences by industrial activity



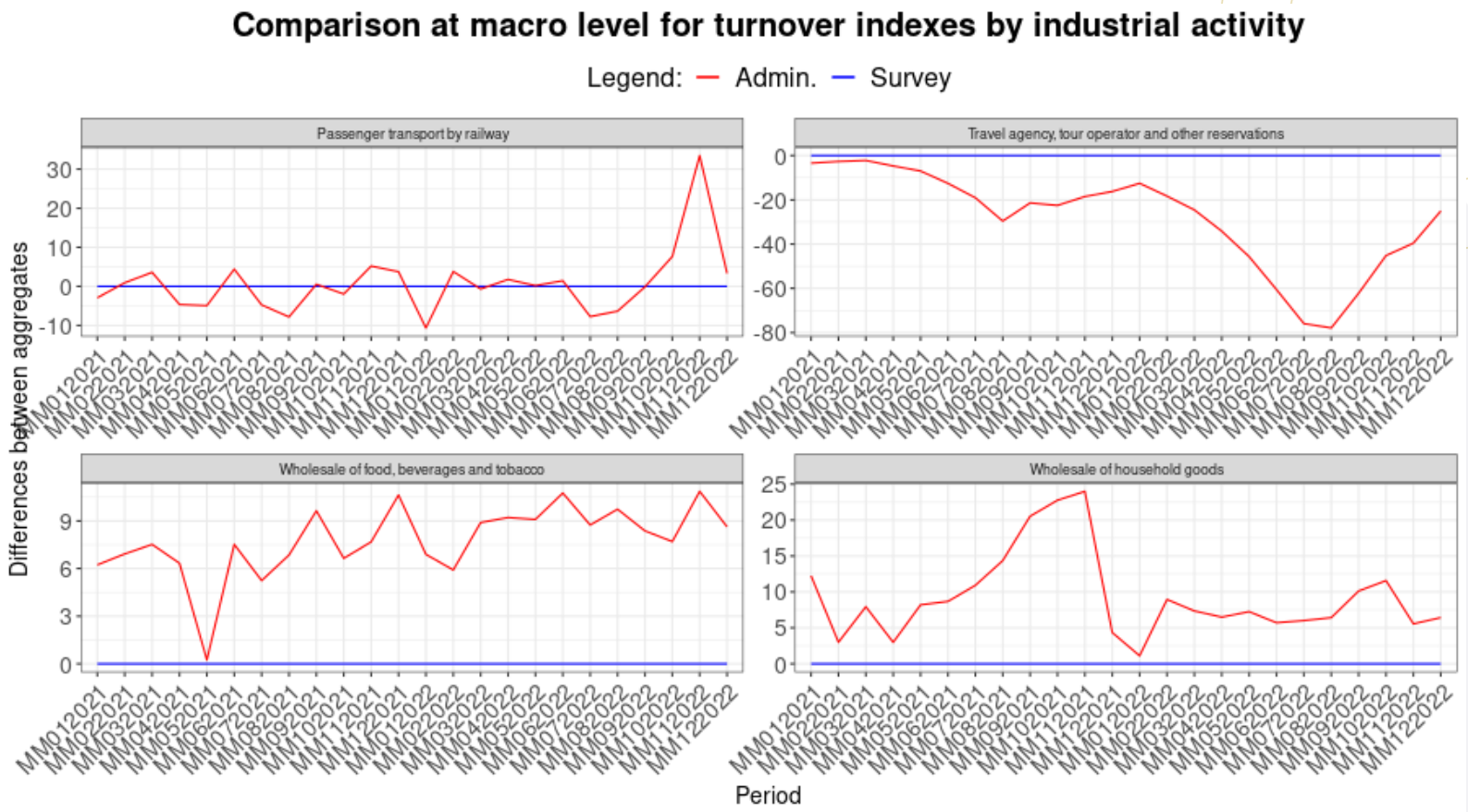
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Data comparison for aggregated-level macrodata: turnover differences



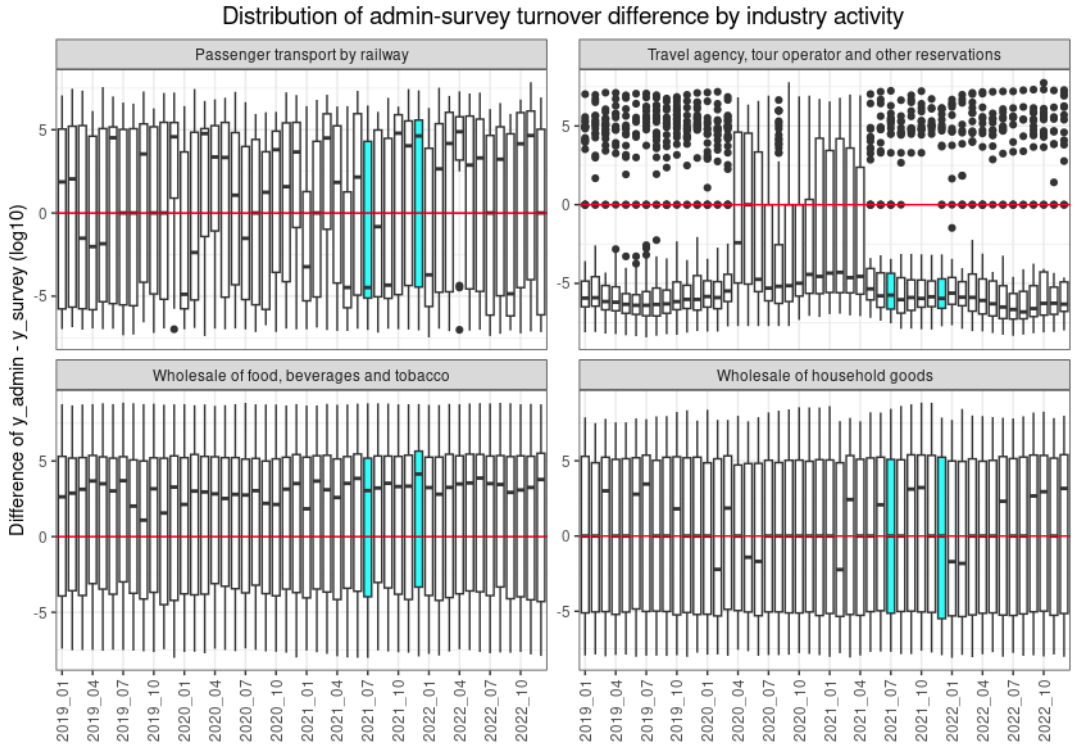
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

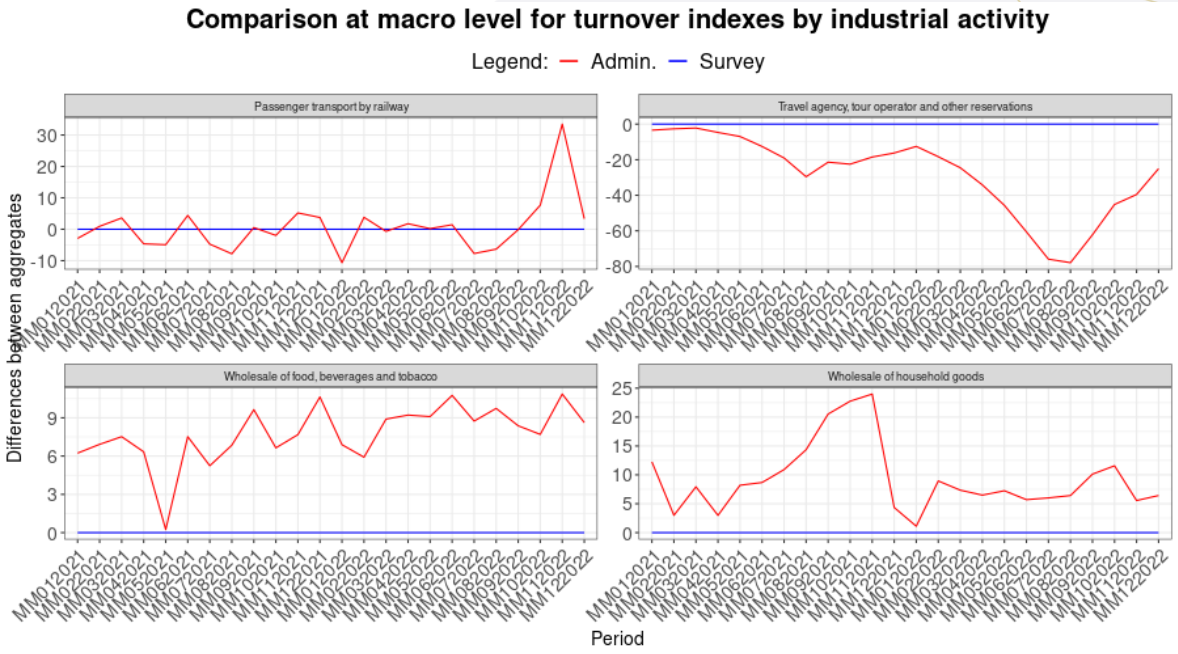
Data comparison for aggregated-level macrodata by industrial activity



The conference is partly financed by the European Union

Behaviour of the **microdata** by activity sector diverges significantly from that of the corresponding **aggregates**!

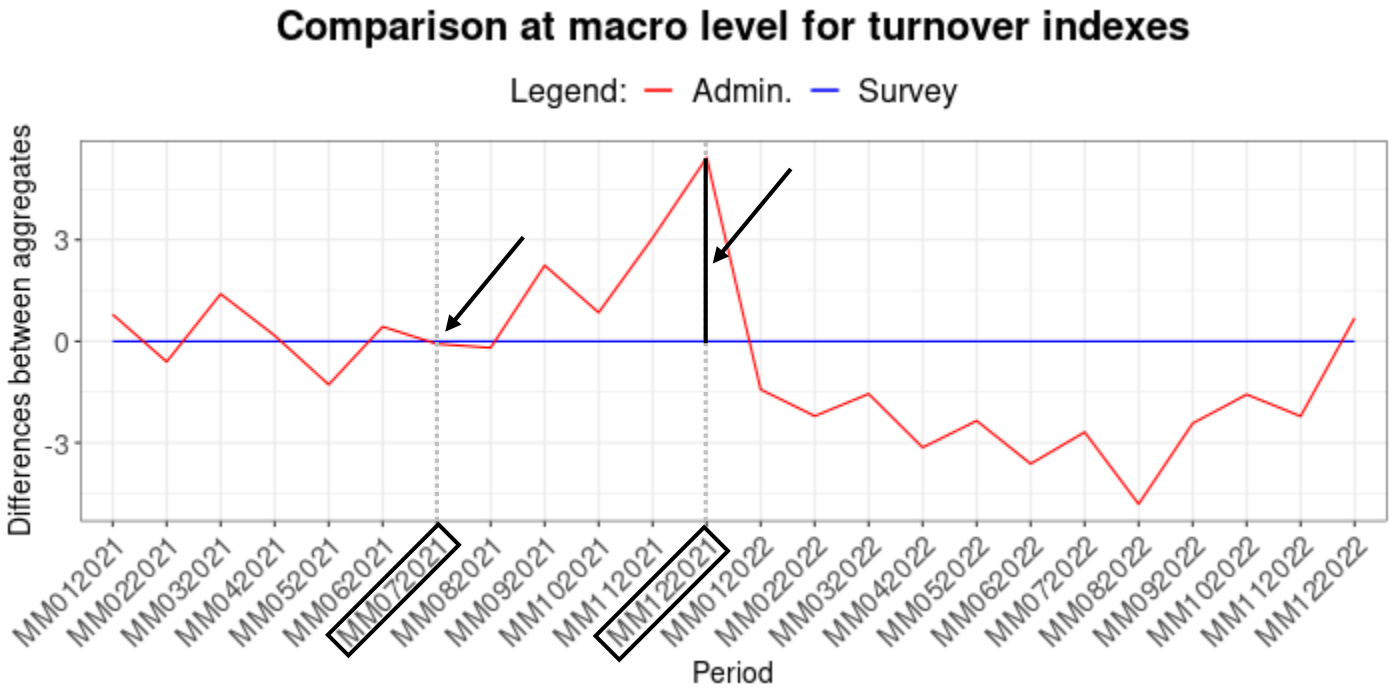




INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

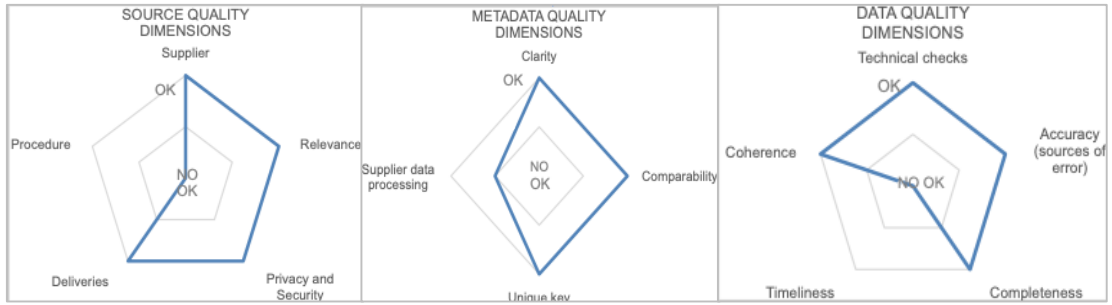
Can we anticipate the performance of a data source with input quality indicators?

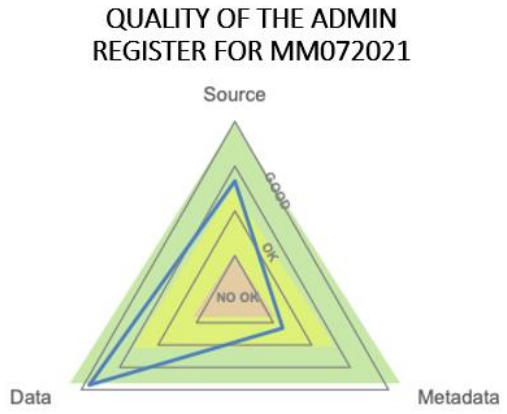
Examining July and December 2021 in particular…



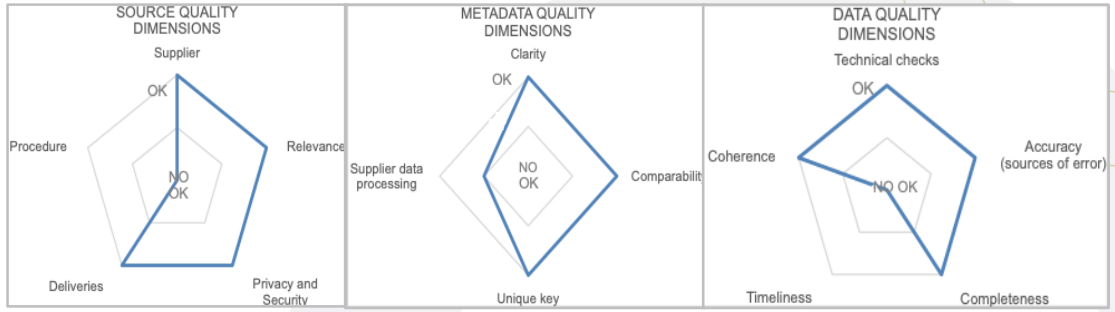
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

July 2021





December 2021



Essentially indistinguishable despite the dissimilarities in the aggregated data!

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Overall conclusions

- The differences observed in the exploration of the input quality **cautiously discourage** the use of the administrative values by mere substitution as the statistical values.
- The **generation process** significantly influences the **quality** of the final results (aggregates). Conversely, the quality of the input (microdata) is determined by both the intended use of the output (aggregates) and the chosen input processing methods.
- There **remains** a significant amount of **work to be done** in order to develop a fast, accurate, efficient, and appropriate method for measuring the quality of administrative data as an input.

Measuring the quality of administrative sources: at macro level with novel indicators and micro level with distributions comparison

Alicia Nieto, Sandra Barragán, Alba Rodríguez,
Soledad Saldaña, David Salgado

S.G. for Methodology and Sampling Design
Statistics Spain

June 2024

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS

2024 ESTORIL - PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union