
Detecting Extreme Numerical Outliers in Trade Data: A Novel Method for Highly Asymmetric Distributions

Andrea Cerasa (European Commission's Joint Research Centre)

andrea.cerasa@ec.europa.eu

Abstract

The accurate detection of extreme numerical outliers in trade data is crucial for effective policy making, anti-fraud measures, and reliable EU-wide statistics. The European Commission relies heavily on the Surveillance database of DG TAXUD, which collects import/export transactions from national authorities. However, large errors in declared values may occur due to data quality issues, and this can severely impact data analyses and lead to incorrect decisions.

The significant skewness observed in the distributions of international trade data poses a major challenge in identifying these extreme numerical outliers, as it can be difficult to distinguish them from the normal values in the right tail of the distribution. Building upon the work described in the scientific article “An adjusted boxplot for skewed distributions” (Hubert- Vandervieren 2008), our proposed method addresses the challenges posed by the highly skewed distributions found in international trade data. In particular, the primary contributions of our proposal are:

1. Enlarging the spectrum of distributions considered in the article to more closely resemble the asymmetric distributions that characterize international trade data. This allows for a more accurate representation of trade data distributions, which are often more skewed than those found in the existing literature.
2. Developing a method for calculating thresholds that identify extreme anomalous numbers of each distribution, rather than adapting a box plot for skewed distributions. This novel approach is specifically designed for detecting extreme numerical outliers in highly asymmetric distributions.

To assess its quality, we will test it on real international trade data provided by DG TAXUD. This practical deployment will enable us to evaluate the effectiveness of our approach in detecting extreme numerical outliers in trade data and contribute to the improvement of data quality checks at DG TAXUD. It will also assess the potential to significantly enhance the reliability of EU-wide statistics, anti-fraud measures, and policy making, as well as facilitate economic operators in their activities.

To assess its quality, we will test it on real international trade data provided by DG TAXUD. This practical deployment will enable us to evaluate the effectiveness of our approach in detecting extreme numerical outliers in trade data and contribute to the improvement of data quality checks at DG TAXUD. It will also assess the potential to significantly enhance the reliability of EU-wide statistics, anti-fraud measures, and policy making, as well as facilitate economic operators in their activities.