UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

UNECE Machine Learning for Official Statistics Workshop 2023

04 April 2023

5-7 June 2023, Geneva Switzerland

Timeliness and Accuracy with Machine Learning Algorithms: Early Estimates of the Industrial Turnover Index

David Salgado, Sandra Barragán, Elena Rosa-Pérez (Statistics Spain)

david.salgado.fernandez@ine.es

Abstract

We use statistical learning algorithms to improve timeliness of the Spanish Industrial Turnover Index. The main idea is to use a gradient boosting algorithm to make a prediction for every single industrial turnover value not yet collected during the data collection, data editing and estimation phases. Regressors are constructed from the historical unit-level time series, current aggregated turnover moments and quantiles, and aggregated values of related industrial surveys. The proposed construction of regressors is strongly based on and motivated by the data validation criteria used by subject-matter experts in standard production.

Accuracy indicators, which does not make use of exchangeability hypotheses for the population units, are also computed so that a quantitative trade-off between accuracy and timeliness can be appraised.

In this way, we can provide increasingly precise early estimates of the index as microdata are collected and edited during regular production.

This mass imputation exercise provides us with a nowcasting proposal which can be readily extended to many similar design-based statistical products based on survey data already in production in multiple statistical offices.