Creator **Madalina Fron** Created **Aug 02, 2023, 13:59** Last updated **Aug 08, 2023, 12:51**

Files

- scope_estimators/sector_model_army.py
- experiments/experiment_SMA_vs_MMA.py
- notebooks/analyze_SMA_vs_MMA.py

Motivation

Based on the Harvard paper (citation below) we train according to a bucket classification approach, more specifically for each sector we train a different regressor. At the moment we train our regressors by bucketing based on scope buckets, but maybe it is easier to make a prediction conditioned on the sector of a company.

Serafeim, George and Velez Caicedo, Gladys, Machine Learning Models for Prediction of Scope 3 Carbon Emissions (June 1, 2022). Harvard Business School Accounting & Management Unit Working Paper No. 22-080, Available at SSRN: https://ssrn.com/abstract=4149874 or http://dx.doi.org/10.2139/ssrn.4149874

Design

To test the behaviour we implemented two variants of the MMA model, the DirectSectorModelArmyEstimator and the SectorModelArmyEstimator. The DSMA estimator takes the sector feature specified in the data point and uses it as the grouping key. This works only if we can assume that the user always specifies the sector when making a prediction.

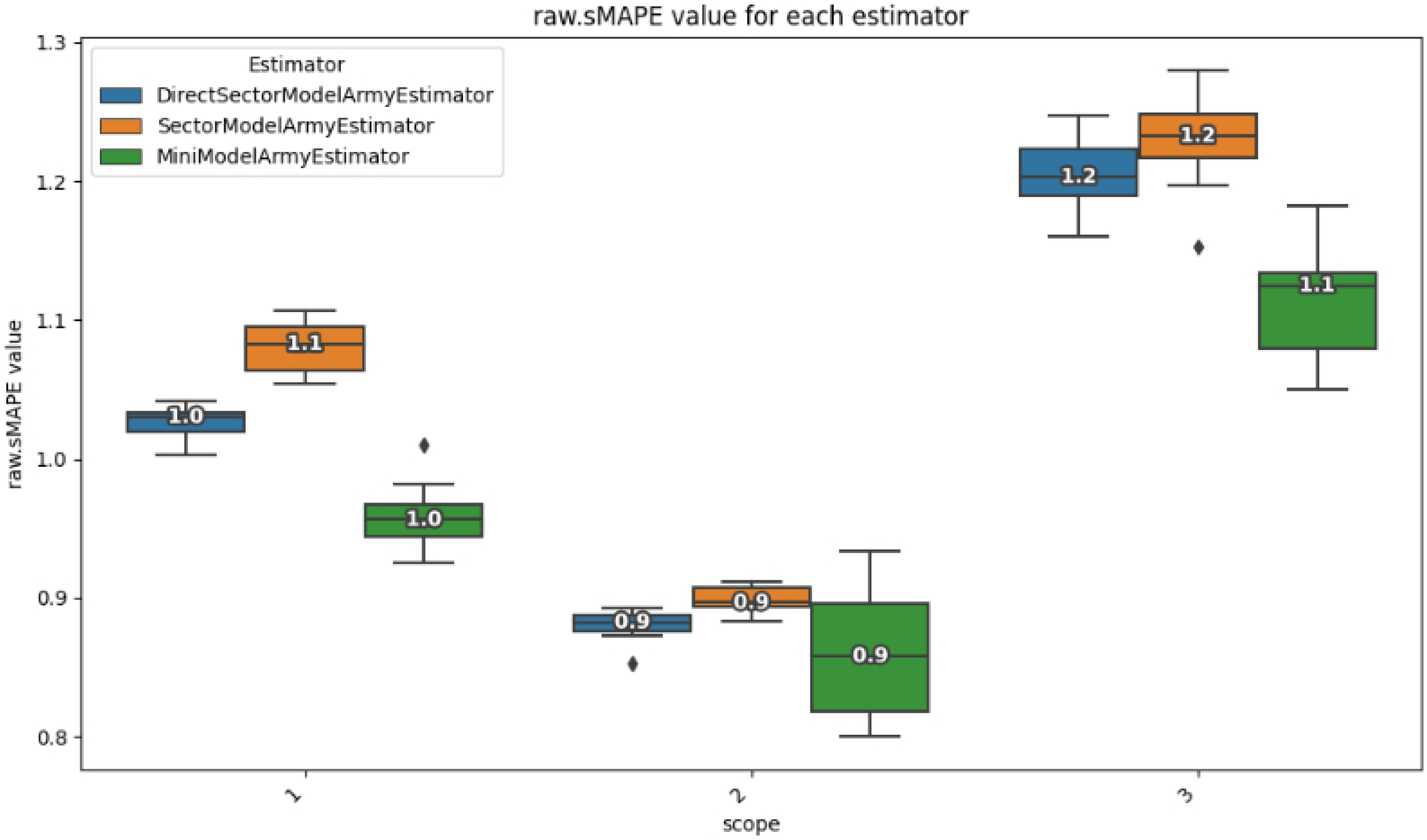
For the case when we cannot assume the provision of the sector feature, we need to find a way to estimate the most probable sector, based on the features that are available in the data point. For this purpose, the SMA estimator trains a classifier which uses the sector feature as target and the other features as independent variables.

We ran the experiment on the full dataset 10 times for all estimator configurations, each configuration using a new set of pipelines and considering all 3 scopes.

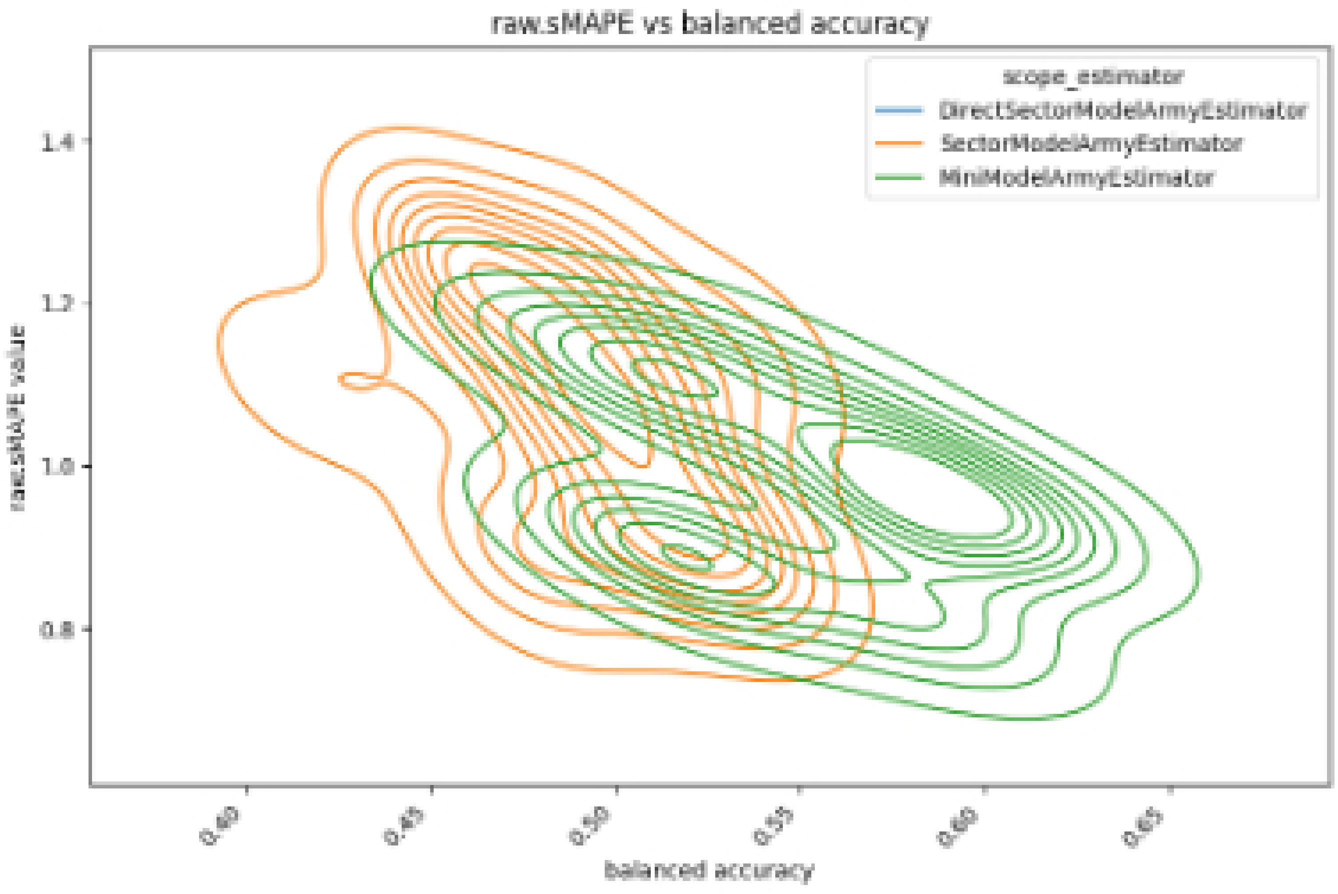
ⓘ The comparison models do not perfectly with the Harvard approach. They are just inspired by it. The paper only tried to predict scope 3, and they used a different set of features.

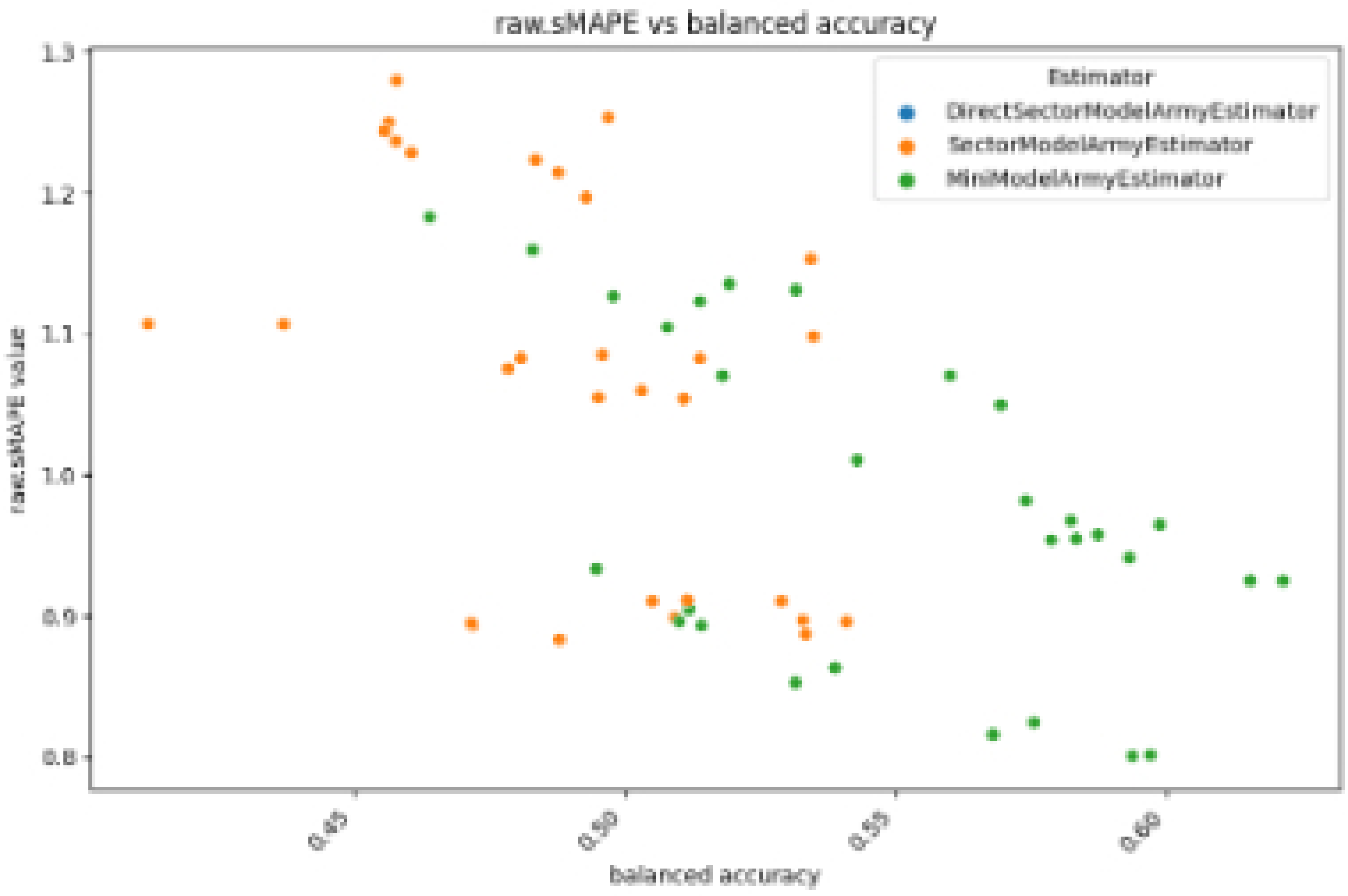
Results and insights

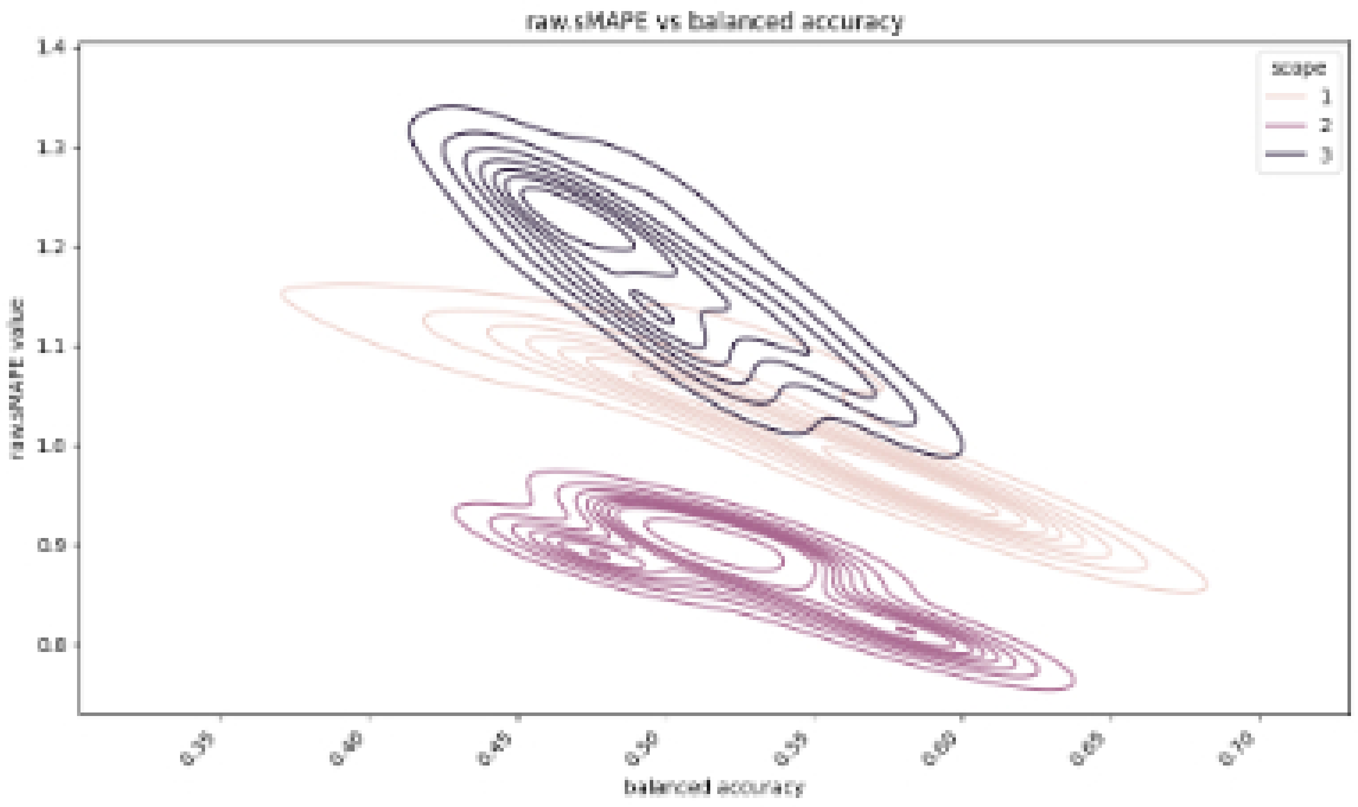
For all scopes, the sMAPE for the MMA estimator has the lowest medians. For scopes 2 and 3, the variance of the MMA is greater that the other ones, but this might happen because it's fundamentally harder to predict the right scope bin.

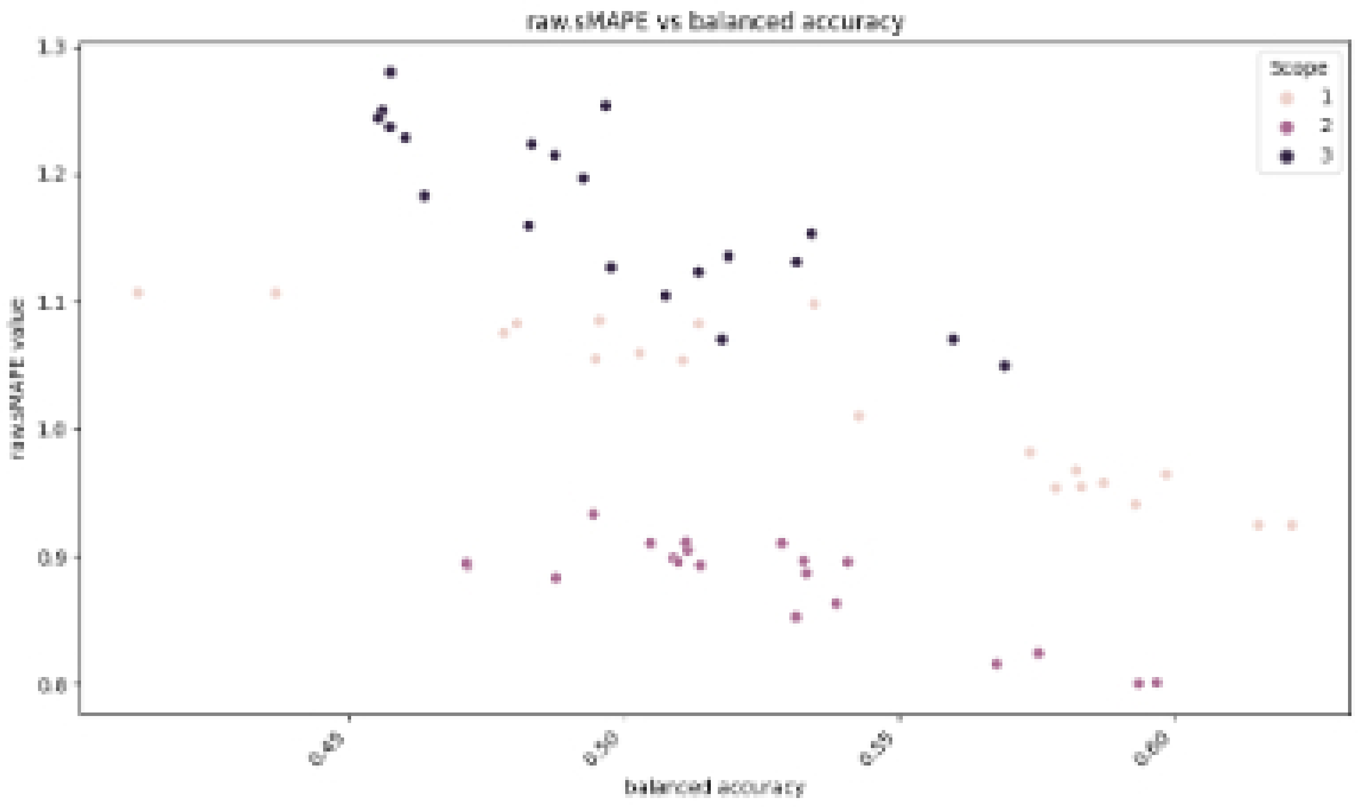


The plots below show a strong relationship between the balanced accuracyand the sMAPEvalue both for SMA and MMA : increasing accuracyleads to decreasing sMAPE, which is desired. The MMA estimator still leads to higher accuracy. The DSMA estimator is not shown in these plots becausethe accuracyscore cannot be calculated for its design.









Decision

We will continue using the original MMA. It was interesting to see how the Harvard approach compares to it and how it fits our data. The results were not that far off afterall, the difference in performance being of maximum 0.1.