Creator **Olusanmi Hundogan** Created **Dec 18, 2023, 17:21** Last updated **Mar 12, 2024, 14:42**

Files

- notebooks/analyze_weighted_voting_vs_even_voting.py
- experiments/experiment_weighted_voting_vs_even_voting.py
- scope_estimators/mini_model_army.py
- scope_estimators/mma/regressor.py

Motivation

The current bucket regressor uses a Voting Mechanism with several different sub-models to produce an estimation for the scopes. To weight the individual sub-models, the Mini Model Army estimator calculates the performance score of each individual sub model within each bucket. The chosen performance metric for the models was smape. In this experiment, we were interested in the effectiveness of the weighting mechanism. Thereby, testing this current bucket regressor against two other options: a regressor which uses alternative cross-value metric, namely the median absolute error, and another regressor which assigns even weights to all models within a bucket.

Design

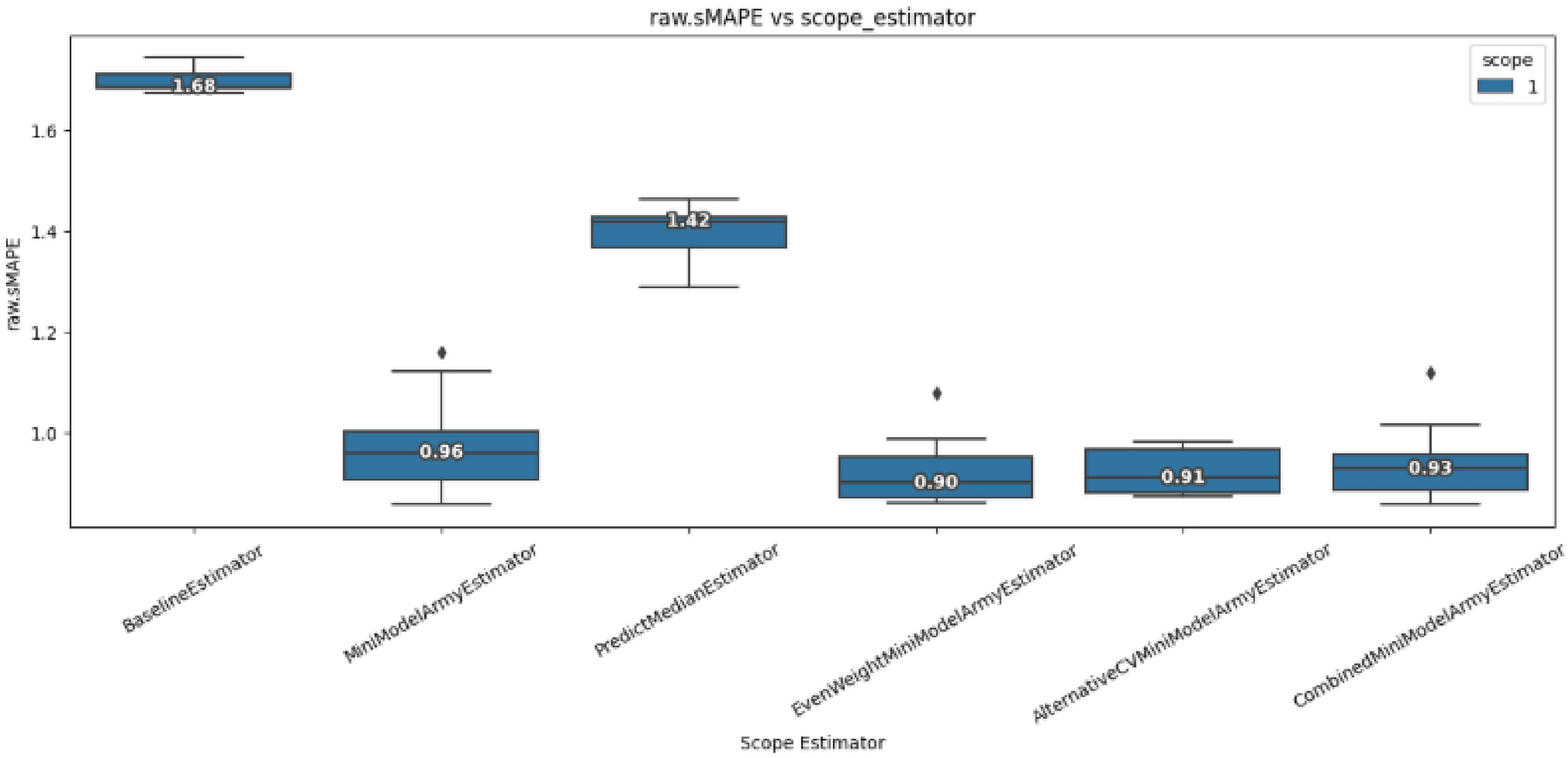
The experiment has 10 repetitions and each repetition uses a new small dataset configuration (fraction of 0.1). It is based only on scope 1 data. A repetition loops through the following estimators:

- "EvenWeightMiniModelArmyEstimator": This estimator uses even weights for every sub-model;
- "AlternativeCVMiniModelArmyEstimator": This estimator uses the median absolute error as cross-value metric instead of sMAPE;
- "CombinedMiniModelArmyEstimator": This estimator uses even weights **and** a different cv-metric (MAE).
- "MiniModelArmyEstimator": This estimator uses the smart weights for the voting regressor;
- "BaselineEstimator": This estimator only predicts randomly selected values of the scope values it trained on.
- "PredictMedianEstimator": This estimator always predicts the median of the scope values it trained on.

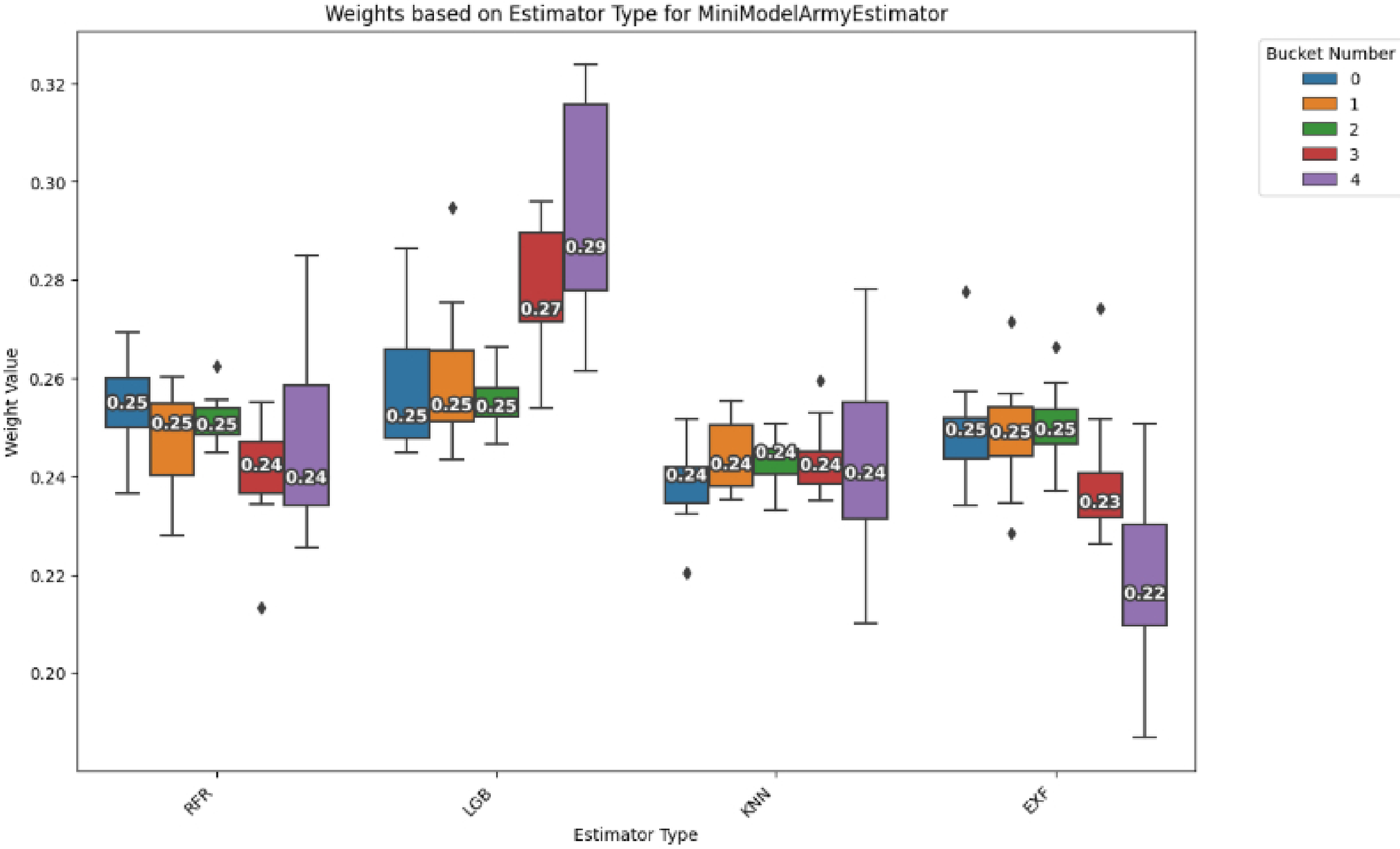
Each repetition creates a new model pipeline for each estimator and fits it according to the corresponding regressor. Then the bucket specifics are saved as results and analysed below.

Results and Insight

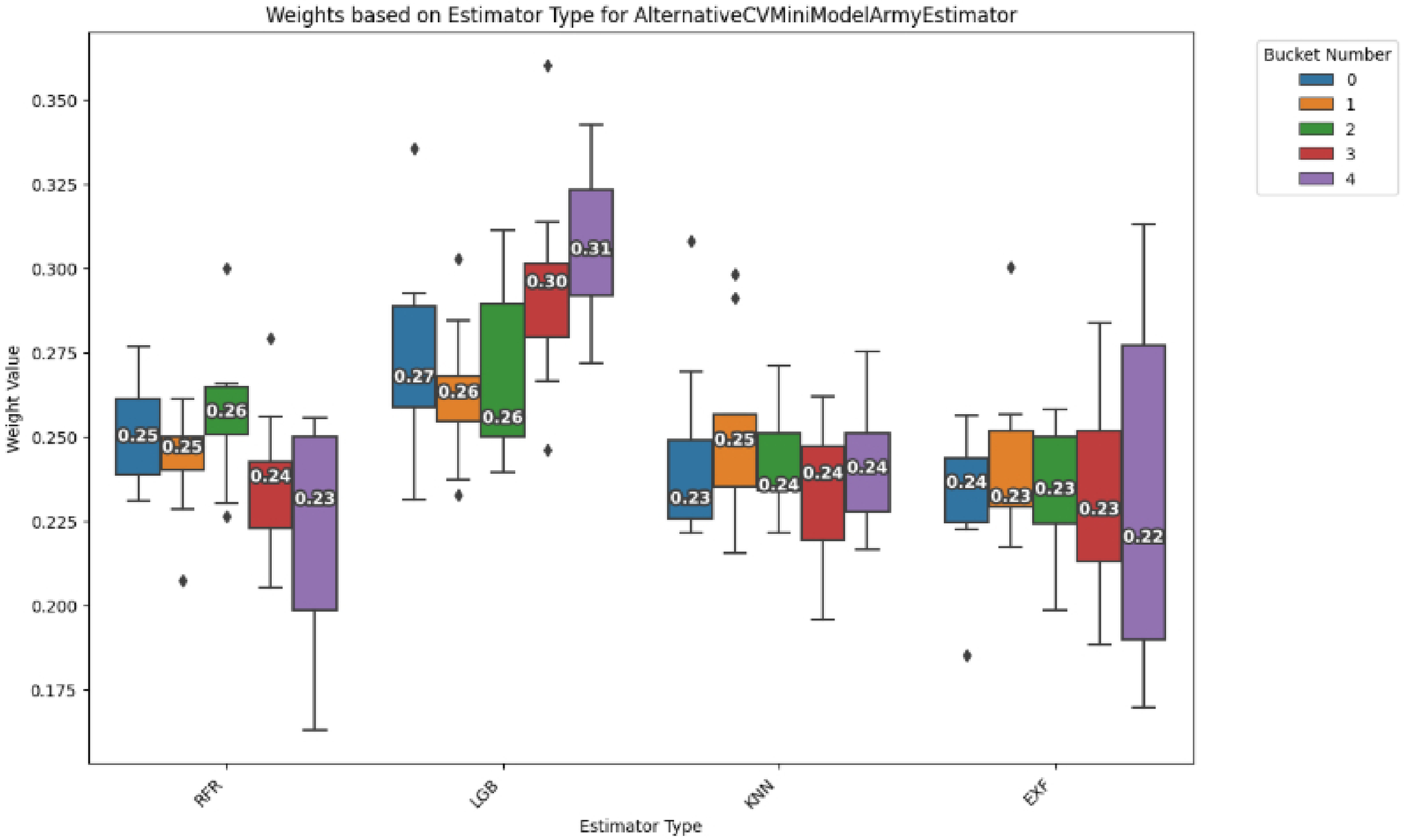
The overall results show that the weighted solution performs worse that then even weighted configuration. In fact, even weighting performs best among all other configurations. Furthermore, the application of a different metric for cross validation improves the weighted solution as well. However, using even weighting and a an alternative cross validation metric such as MAE does not compound to an improvement but rather to a slight decrease in performance. Interestingly, there were no performance outliers for the alternative CV method.



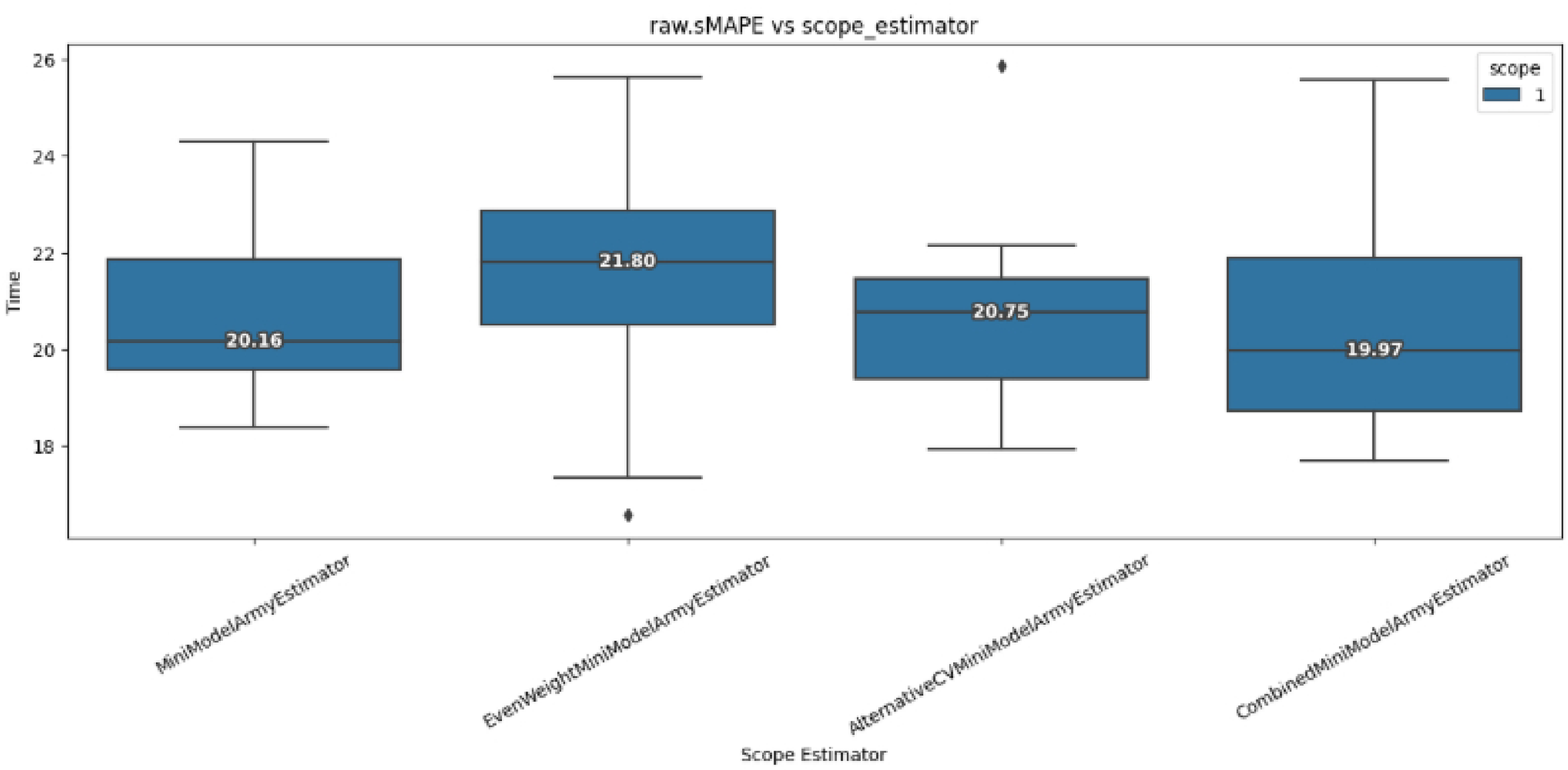
For the MMA Estimator, which uses smart weights for the regressor and sMAPE as cv-metric, LGB has the highest score (weight) for voting in each bucket, as expected. The greatest difference in weight range can be seen in the last bucket because there are the most difficult numbers to predict. (The buckets are divided with percentiles and the last percentiles have the widest spread as the scope values are heavily right skewed.)



We can see an even more accentuated result when using the alternative cv-metric, MAE.



With regards to time we expected the even weighted estimator to be more time efficient. However, we see that the even weighted case is the slowest estimator and the combined version is the fastest although both lack the sub model cross validation mechanism. This inconsistency might stem from the CV mechanism not having any substantial effect on the time estimate. We assume that the results here are based on random fluctuations in processing time.



Decision

Update : 29.12.2023

Clearly, the current approach performs worse than the other options explored in this experiment. The other configurations all performed similarly with the even weight configuration emerging as the best performing option.

We conclude for now that we should switch to use even weight, because this configuration reduces unnecessary complexity within the model and improves its performance visibly. Alternatively, we can choose to use a different cv-metric such as MAE but that would keep most of the initial complexity.

In the future we will also observe compare whether Voting regression in general competes with other Ensemble methods such as Stacking.

Update : 12.03.2024

With newly available data the results turned out in favor of AlternativeCV.