An Phríomh-Oifig Staidrimh
Central Statistics Office

Model Assisted Estimation

Sorcha O'Callaghan
Central Statistics Office, Ireland

Background

- Falling response rates mean we want to incorporate more admin data
- Have obtained new datasets from Irish Revenue
- Use auxiliary variables to train ML models to predict variables
- Use Model Assisted Estimator as proposed by Breidt & Opsomer
- https://projecteuclid.org/journals/statistical-science/volume-32/issue-2/Model-Assisted-Survey-Estimation-with-Modern-Prediction-Techniques/10.1214/16-STS589.full
- Analysis by by Dagdoug, Goga and Haziza on MAE using Random Forest (https://arxiv.org/pdf/2002.09736.pdf)

Issues

- Determining a suitable model – working with RF, KNN, SVR
- Increasingly need to incorporate non-probabilistic data sources

What is the Model Assisted Estimator

Difference Estimator:

$$Diff(y, m) = \sum_{k \epsilon U} m(x_k) + \sum_{k \epsilon s} \frac{y_{k-m(x_k)}}{\pi_k}$$

Model Assisted Estimator:

$$Diff(y, \hat{m}) = \sum_{k \epsilon U} \hat{m}(x_k) + \sum_{k \epsilon s} \frac{y_{k-\hat{m}(x_k)}}{\pi_k}$$

An Phríomh-Oifig Staidrimh
Central Statistics Office

Feedback

Interaction with the group

- Quarterly reports
- Feedback on bias, variance, accuracy issues
- What types of model are working best for people?

title

An Phríomh-Oifig Staidrimh
Central Statistics Office