

Comments on "Imputation Techniques for the Nationality of Foreign Shareholders in Italian Firms" by Carboni A., and A. Moro

Gianluca Cubadda

Università di Roma "Tor Vergata"

October 21, 2019

- The empirical problem is the imputation of the nationality of shareholders within the FDI survey.
- The authors propose a procedure, based on text mining and classification ML techniques, which exploits the name of corporations only as input data.
- The algorithm is trained using the database Orbis, where the nationality of the single enterprise is known.
- According to an extensive out of sample exercise, the new procedure outperforms the current practice in terms of correct classification.
- Lots of subtle questions, mostly related to data selection and cleaning, need to be solved before applying the procedure. Prior human knowledge is essential for the success of ML strategies.

- Singular value decomposition of predictors is an unsupervised dimension reduction technique. Why not relying on *supervised* PCR (Bair *et al.*, 2006, JASA) or SVD (Li *et al.*, 2016, JMA)?
- Recent advances on regularized regression methods convincingly argue that the *adaptive* lasso (Zou, JASA, 2006) outperforms in prediction the ordinary one. Such approach could be used to regularize the (multinomial) logit models.
- In order to assess the merits of the proposed procedure, Table 3 should be compared with the subsequent ones by some formal statistical procedure.
- Why not a 5 or 10 fold cross-validation instead of a unique validation set?