

## General discussion: Method (A)

---

Babet: Wage imputation by NLP deep learning

- existing system: *ad hoc*, not adequately tested
- well-known parametric model: high explainability, yet...
- predictive ML algorithms: not without drawbacks, but still...

Preising: missing data in linear dynamic panel regression

- missing items both in  $NT$ -vector  $y$  and  $NT \times K$ -matrix  $X$
- Gibbs sampler incorporating observed items is more efficient

Jentoft: data editing for machine learning

- filtered training set may be necessary, but how?
- what can balancing feature distribution in training set achieve?

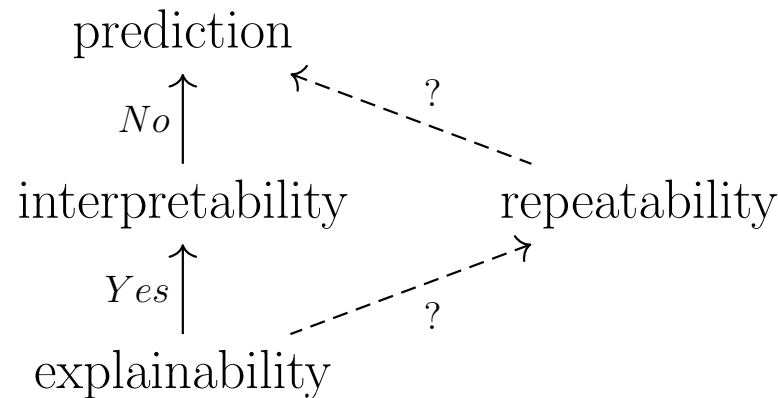
Leather: RBEIS software

- restricted hot-deck reduces variance by imputation otherwise

## Some questions and remarks

---

Babet: Wage imputation by NLP deep learning



Preising: missing data in linear dynamic panel regression

- small-data simulation setting:  $N = 10$ ,  $T = 5$ ?
- normality of complete data, simple missing-data assumption...

Jentoft: data editing for machine learning

- lacks ML approach that models outcome and error jointly?
- confounding feature imbalance and informative sample selection?

Leather: RBEIS software

- efficiency gain over unrestricted hot-deck can be measured...