# Discussion on


*Roberto Rocci*

# Paper

**Context** Data Quality Management (DQM)

**Aim** Identify potential outliers

ex Y ≡ # cards issued by bank *i*, province *p*, semester *t*

**Method** 1) Estimate the conditional distribution of Y|x through random forest quantile regression

2) Data out of a quantile-based interval is labelled as potential outlier

# What I like

- Paper

  interesting, well written

- Central idea

  **quantile regression** (estimates of conditional quantile functions are more robust than estimates of conditional mean functions) **forest** (very flexible way to model relations between Y and **x**)

- (Statistical) Machine Learning

  does it work? Let's try and compare

# What I do not like

Lack of an interpretable model.

Domain information cannot be used to check/validate the model

# Suggestions/Questions

- # of predictors seems to be quite small.

- Computational complexity?

- How sensitive are the results to the setting of tuning parameters (# trees, # of variables at each split, replace = ?)?