36th IARIW General Conference

Paper Prepared for the 36th IARIW General Conference, Oslo, Norway, August 23-27, 2021

Hedonic Imputation with Tree-based Design Approaches

Shipei Zeng

Linear hedonic regression is commonly utilised to estimate missing prices in cases where there is product entry and exit, or product "churn", but the linear assumption of prices in product characteristics is dubious. Actual consumer purchase patterns show that product characteristics are not perfectly substitutable so that the prediction capacity of linear models is challenged. I consider alternative estimations of hedonic prices by introducing tree-based machine learning models that are highly recommended for prediction accuracy. Particular attention is paid to the micro-economic explanation of tree-based models. A tree decision structure is compatible with consumer preferences when product characteristics are complements. Model performance metrics from (electronic-point-of-sale) scanner data confirm prediction accuracy gains from the appropriate model selection that follows consumer behaviour foundation. I find that random forests are the best fitted model with largest $\bar{R}^2$-type measures among a series of models. Price indexes with random forests display correct predictions that are robust in the single, double and full imputations. The variable importance estimated for product characteristics is consistent with the actual coefficients of hedonic functions in price simulation. It is advisable that tree-based decision approaches, especially random forests, can be effectively employed for unmatched products in hedonic imputation due to their prediction accuracy and compatibility with consumer utility types.