Automated Variable Selection on German Real Estate Data Using Structured Additive Distributional Regression

Julian Granna, Stefan Lang

November 22, 2022

Abstract

Modeling real estate prices in the context of hedonic models typically involves fitting a Generalized Additive Model, where only the mean of a (lognormal) distribution is regressed on a set of covariates without taking other parameters of the distribution into account. Moreover, the choice of variables to include in the model involves costly, sometimes infeasible, computations, even on powerful computers. This is especially true for the selection of relevant, possibly complex interaction terms.

We apply a novel backfitting algorithm in the context of a structured additive regression model that enables us to select and model all distributional parameters of an appropriate distribution. Using a large German dataset on rental prices, we first select an appropriate distribution from a set of candidates. In a second step, we are able to efficiently, automatically select covariates along with relevant interactions between them allowing for e. g. spatial heterogeneity.