Predicting Housing Prices in Denmark
A Machine Learning Approach

Group 23: He SHI, Mingjie XING, Qi ZHANG
Department of Economics, University of Copenhagen Summer 2023

August 15, 2023

1. What is your research question?

   This project is intended to build a predictor of residential housing prices in cities in Denmark with a machine learning hedonic model based on interior features such as house area and location, and exterior features such as income level of the neighbourhood, weather condition and cities' economic development level. The project mainly covers major cities in Denmark, e.g., Copenhagen and Aarhus. The frequency of the data will be monthly. The trained model based on 2023 data will be tested on 2022 and 2021 data.

2. What kind of data are you planning on using? How will you get access to these data?

   - House Prices Data:

     We are going to scrape the individual level house prices from boligsiden, a major real estate broker in Denmark covering housing prices as monthly data of 36 months scale and detailed features of houses in different regions with an open data access. We will scrape the data from `https://www.boligsiden.dk/`. Structural features in data includes area, location, age and owner income. Web-scraping of the data is allowed for academic use. There will be about 10000 house-month-region observations in our dataset.

3. What will your data analysis be like? Will you use machine learning? How?

We are going to train a machine learning model on residential housing prices in Copenhagen in 2023 in each postal coded region. The feature variables include structural features such as house area and age, and external variables include location, average income of the neighbourhood and weather conditions of each city. We are going to test the model with 2022 data and data from other cities such as Aarhus.

We are going to apply the polynomial and LASSO regression on the data to obtain the coefficient of deciding factors on house price in different cities in Denmark and in different temporal sphere.

We will handle the under-fitting and over-fitting problem when applying the machine learning model to get a more precise version of the house price prediction.

We will conduct cross-validation with k-fold method.

4. Have you already identified other papers within this area that you can use in a literature review? If so, name a few and explain what they do in one sentence only.

There are several papers recently on top journals in real estate economics predicting housing prices with machine learning models:
Deppner et al. (2023) applies non-parametric machine learning model with k-fold cross-validation to examine the U.S. commercial real estate appraisal to shrink the deviations between market values and subsequent transaction prices.
Lin et al. (2023) augments the traditional hedonic model with Gradient Boosting Machine algorithm to predict housing prices in Beijing.
Tchuente and Nyawa (2022) uses artificial neural networks, random forest, adaptive bBoosting, gradient boosting and K-nearest neighbours algorithms to estimate real estate prices in French cities.

5. How do you 'contribute' to the literature?

The main contribution of this project includes: 1) creating a comprehensive dataset on Denmark residential real estate market; 2) building a predictor of Denmark housing prices in different cities.

Bibliography

Deppner, J., von Ahlefeldt-Dehn, B., Beracha, E., and Schaefers, W. (2023). Boosting the accuracy of commercial real estate appraisals: An interpretable machine learning approach. *The Journal of Real Estate Finance and Economics*.

Lin, W., Shi, Z., Wang, Y., and Yan, T. H. (2023). Unfolding beijing in a hedonic way. *Computational Economics*, 61(1):317–340.

Tchuente, D. and Nyawa, S. (2022). Real estate price estimation in french cities using geocoding and machine learning. *Annals of Operations Research*, 308(1):571–608.