Take-home Exercise 3:

This handout provides the context, the task, the expectation and the grading criteria of Take-home Exercise 2. Students must review and understand them before getting started with the take-home exercise.

Contents

Setting the Scene

Housing is an essential component of household wealth worldwide. Buying a housing has always been a major investment for most people. The price of housing is affected by many factors. Some of them are global in nature such as the general economy of a country or inflation rate. Others can be more specific to the properties themselves. These factors can be further divided to structural and locational factors. Structural factors are variables related to the property themselves such as the size, fitting, and tenure of the property. Locational factors are variables related to the neighbourhood of the properties such as proximity to childcare centre, public transport service and shopping centre.

Hedonic pricing model is used to examine the effect of housing factors as discussed above on the price. Conventional, this model was built by using **Ordinary Least Square (OLS)** method. However, this method failed to take into consideration that spatial autocorrelation and spatial heterogeneity exist in geographic data sets such as housing transactions. With the existence of spatial autocorrelation, the OLS estimation of hedonic pricing models could lead to biased, inconsistent, or inefficient results (Anselin 1998). In view of this limitation, **Geographical Weighted Regression (GWR)** was introduced for calibrating hedonic price model for housing.

The Task

In this take-home exercise, you are tasked to build hedonic pricing models to explain factors affecting the resale prices of public housing in Singapore. The hedonic price models must be built by using appropriate GWR methods.

The Data

For the purpose of this take-home exercise, `Resale Flat Prices` provided by Data.gov.sg should be used as the core data set. The study should focus on four-room flat and transaction period should be from 1st January 2019 to 30th September 2020.

Below is a list of recommended independent variables to consider. However, students are free to include other appropriate independent variables.

- Structural factors
  - Area of the unit
  - Floor level
  - Remaining lease
  - Main Upgrading Program (MUP) completed (optional)
- Locational factors
  - Proxomity to CBD
  - Proximity to eldercare
  - Proximity to foodcourt/hawker centres
  - Proximity to MRT
  - Proximity to park
  - Proximity to good primary school
  - Proximity to shopping mall
  - Proximity to supermarket
  - Numbers of kindergartens within 350m
  - Numbers of childcare centres within 350m
  - Numbers of bus stop within 350m

- Numbers of primary school within 1km

Grading Criteria

This exercise will be graded by using the following criteria:

- **Geospatial Data Wrangling (20 marks):** This is an important aspect of geospatial analytics. You will be assessed on your ability to employ appropriate R functions from various R packages specifically designed for modern data science such as readr, tidyverse (tidyr, dplyr, ggplot2), sf, sp just to mention a few of them, to perform the entire geospatial data wrangling processes. This is not limited to data import, data extraction, data cleaning and data transformation. Besides assessing your ability to use the R functions, this criterion also includes your ability to derive appropriate variables to meet the analysis needs.
- **Geospatial Analysis and Modelling (30 marks):** In this exercise, you are expected to use appropriate statistical analysis and GWR functions introduced in class to calibrate hedonic price models. The focus of this criterion should go beyond discussing the modelling results, but include Exploratory Data Analysis, multivariate analysis analysis for detecting multicollinearity, and spatial autocorrelation test, just to name a few of them.
- **Geovisualisation (20 marks):** In this section, you will be assessed on your ability to communicate the complex spatial statistics results in business friendly visual representations. This course is geospatial centric, hence, it is important for you to demonstrate your competency in using appropriate geovisualisation techniques to reveal and communicate the findings of your analysis.
- **Reproducibility (20 marks):** This is an important learning outcome of this exercise. You will be assessed on your ability to provide a comprehensive documentation of the analysis procedures in the form of code chunks of RMarkdown. It is important to note that it is not enough by merely providing the code chunk without any explanation on the purpose and R function(s) used.
- **Bonus (10 marks):** Demonstrate your ability to employ methods beyond what you had learned in class to gain insights from the data. The methods used must be geospatial in nature. (Hint: Mixed Geographically Weighted Regression Model)

Submission Instructions

- The write-up of the take-home exercise must be in **distill** or **blogdown** format. You are required to publish the write-up on **Netlify**.
- The R project of the take-home exercise must be pushed onto your Github repository.
- You are required to provide the links to Netlify service of the take-home exercise write-up and github repository on eLearn.

Due Date

7th November 2021 (Sunday), 11.59pm (midnight).

Useful R packages

- httpr: Useful tools for working with HTTP organised by HTTP verbs (GET(), POST(), etc). Configuration functions make it easy to control additional request components (authenticate(), add_headers() and so on).
- onemapsgapi: An R wrapper for the 'OneMap.Sg' API https://docs.onemap.sg/. Functions help users query data from the API and return raw JSON data in "tidy" formats. Support is also available for users to retrieve data from multiple API calls and integrate results into single dataframes, without needing to clean and merge the data themselves. This package is best suited for users who would like to perform analyses with Singapore's spatial data without having to perform excessive data cleaning. For more detail, refer to this link.

Q & A

Peer Learning

AVULA NIHARIKA

CHIN SHAN PEI SARAH

DARRYL KWOK BING HENG

GOH JIA XIN GENICE

JEROME GOH TING CHUAN

KWEK YI CHEN

MEGAN SIM TZE YEN

NGAH XIN YAN

NOR AISYAH BINTE AJIT

TEO JUN PENG

TOH JUN LONG

WONG WEI LING

XU PENGTAI

YU YILING