PILOT VERSION OF CODE FOR THE WEB-BASED TOOL FOR HEDONIC PRICING

Description of the web-based tool

WP 4
SRC
Gustav Engström

Milestone MS29

SEVENTH FRAMEWORK PROGRAMME

Gustav Engström and Erik Andersson
SRC, Sweden

Jun18th 2015

TABLE OF CONTENTS

1 INTRODUCTION

This milestone outlines a novel method for data collection for hedonic pricing analysis. The task was to develop a web-based tool with an underlying real estate market database that merges web-based advertisements with real estate transaction prices. Such information can help authorities and urban planners to acquire information on real estate market participants' preferences for various urban qualities, including access to green areas. This tool will be used to develop models of hedonic pricing and to do mapping analysis of relationships and distances to different qualitative aspects of green space (see GREEN SURGE D4.2, D4.3; Engström, *forthcoming*). The software has been tested in Malmö and Stockholm, Sweden. For further guidance in how to use the tool and methods for analysing the resultant database we refer to Engström (forthcoming) and GREEN SURGE D4.2.

2 DESCRIPTION OF THE WEB-BASED TOOL

The task of creating a web-based tool for hedonic pricing is now completed. The web-based tool has involved collecting and merging data from three sources i) data from Mäklarstatistik AB ii) data collected from Sweden's largest housing advertisement website Hemnet and iii) Google geocoding services.

We now have a database consisting of sales on the apartment, house and property market, covering approximately 70 percent of all sales made through real estate brokers in Sweden. The database is continually updated with new homes from the above sources, which have been merged into a joint database. The process can be described in four steps:

1. ***Through a contract with Mäklarstatistik AB we have continuously bought monthly nationwide data containing, final sales prices of homes and a multitude of characteristics. According Mäklarstatistik the database covers approximately 70 percent of all housing sales made through brokers in Sweden.***
2. ***The observations are geocoded using Google´s geocoding api in order to correct faulty location data which sometimes is present in the data from Mäklarstatistik AB.***
3. ***Via a contract with a web development company (Stormfors AB), we have also developed a so-called web spider. Web spider is a program that continuously scans the internet and collects specific information. In our case, the spider is programmed so that it searches of websites containing housing data mainly from Hemnet in order to collect information about homes that are up for sale. These data are continuously down loaded to our database.***
4. ***Finally, the information from steps 1) and 2) together are merged into a single database. This is done by means of an algorithm programmed to identify the housing that has been up for sale for example on Hemnet and then reported as sold by Mäklarstatistik. By controlling certain key attributes in both data sets, we can ensure that homes sold reported in step 1) are merged with the right property advertisement in step 2).***

In the first dataset, collected from Mäklarstatistik, we have access to final prices while other information about the sold dwellings (e.g. monthly fee, number of rooms, information on the balcony, etc.) are often omitted or inadequately completed. According to Mäklarstatistik it this usually depends on individual brokers having little incentive to accurately report this type of information. The data can thus be controlled and corrected by geocoding address data in order to make sure the geographic coordinates are correct. The information from the housing ads on the other hand (retrieved using the web spider) contains detailed information about the attributes of the sold homes. Here the brokers have large incentives to be careful when reporting since the purpose of housing advertisements is to attract as many buyers as possible. Thus, careful descriptions of all attributes that may be of interest to buyers are reported. The final sale price however, is rarely reported, since the ad is instead removed after the sale has taken place. By pooling the data from these two sources in step 3), this creates a detailed picture of the housing market in Sweden. We estimate that on average we manage to supplement the lack of data from Mäklarstatistik with a corresponding property advertisement in approximately 80-90 percent of cases. As an example, based on an excerpt from the database for the period 1 May 2012 to 31 December 2012 for Stockholm County we registered 22,157 sales reported to Mäklarstatistik. Among these items so the algorithm in step 3) managed to identify 19,385 property listings as-

sociated with these sales. Thus we supplement Mäklarstatistiks data with additional information from the housing classifieds in about 87 percent of cases during the period. An example of an extract from the merged database can be seen in Table 1.

Access to the database is via secure connection over the internet. The downloaded data set can then easily be read in most statistical software. The database is stored in a so-called cloud that is administered by Google. As administrators of the dataset, we are obliged, under contract with Mäklarstatistik, to restrict access to the data to *personal use* only and prohibit commercial exploitation or redistribution of the dataset to others. Within these bounds we welcome independent researchers to apply for permission to use the material in their own research. The idea is that the material should contribute as much as possible to improved knowledge for urban planning.

Table 1. Example excerpt from the compiled data base

Contract date	2012-05-01	Living area	35
Move in	2012-05-21	Number of rooms	1
Final price	1520000	Elevator	TRUE
Type	housing cooperative	Build year	1959
Lkf code	18401	Monthly fee	1459
Street name	Djurstigen	Ad publication	2012-04-20
Street nr.	20	Ad archived	2012-05-10
Post code	17071	Real estate agent (code)	-209805830
City	Solna	Real estate company	Notar AB
Coord.(lat,long)	59.3808629976,18.0395714077	Ad days	12
Balcony	FALSE	Ad price	1350000
Floor number	5	Ad hits	1430
Number of floors	6		
Description	There is certainly nothing wrong with this studio apartments as such, it is both pleasant and nice. The panoramic views of Lidingö and Little Värtan's water, really does not make things worse. Location near subway, University / downtown		

Although the project so far has had a successful outcome, there are several opportunities for improvement. Such an improvement would be to expand the dataset with data from other sources such as Metria. This would greatly improve our coverage when we get access to the sales that do not get reported to Mäklarstatistik. Another improvement would be to expand the range of web spider including searches from Booli.

The code for the web-based tool has been programmed in Python and Mysql and can be found in a zip file labelled code_web_based_tool_ms29.zip.

3 INTERNET SOURCES

Svensk Mäklarstatistik AB, see http://www.maklarstatistik.se/

Hemnet, see http://www.hemnet.se/

Google's geocoding service, see https://developers.google.com/maps/documentation/geocoding/

4 LITERATURE

Engström, G. Will proximity to a public park increase the value of your home? Some evidence from the city of Malmö, Sweden. *Manuscript.*

LIST OF PARTICIPANTS

No.	Participant legal name (and short name)	Country	Organisation type
1	Stockholm Resilience Centre, SRC	Sweden	Research institute