REAL ESTATE HEATMAPS FOR MALMÖ, SWEDEN

Maps based on the tool developed within Task 4.1

WP 4
SRC
Gustav Engström

Milestone MS33

Erik Andersson and Gustav Engström
SRC, Sweden

Draft 1 • Jun18th 2015

SEVENTH FRAMEWORK PROGRAMME

TABLE OF CONTENTS

1 INTRODUCTION

A heatmap is a visualization used to depict the intensity of data at geographical points. A heatmap produce a coloured overlay on top of a map, in this case a city map of Malmö. Generally, areas of higher intensity are coloured red, and areas of lower intensity appear blue or green. The data used in this GREEN SURGE milestone are sales prices for real estate properties in Malmö, Sweden, and the values are outcomes of a hedonic pricing analysis. Since data entry, on part of the real estate agents, obviously is prone to the human error, we decided to carry out a number of data cuts in order to minimise the likelihood of errors in our data set. First, we remove all transactions where the data is either missing or extreme. For the final sale and list price this involved removing all transactions with a sale or list price below 10000 SEK (1400$) or above 10000000 SEK (which we considered to be very unlikely). We also removed all observations where the provided living area is zero, negative or missing, and where the monthly fee is negative or missing. Finally, we also decided to also check for internal inconsistencies in the geographical information provided by the real estate agents i.e. we wanted to test whether the address information entered by the real estate agents also matched the coordinates they provided, for each respective transaction.

2 HEATMAPS

The data used in this paper consists of real estate transactions for both ownership apartments in cooperative housing associations and private housing, in the municipality of Malmö and covers the period October 2010 to April 2015. Malmö municipal with a land area equal to 158.39 km2 and a population size of around 300 thousand is located in the south of Sweden, with easy access to Copenhagen via a bridge and tunnel connecting Sweden to Denmark. The transaction data includes information on e.g. sales price, list price, contract date, living area, floor number, monthly fees, number of rooms and two types of geographical information both in terms of an address and county but also as in terms of geographical coordinates (latitude and longitude). The data set was attained from Svensk Mäklarstatistik which together with Statistics Sweden collects the data from the majority of all real estate agent firms in Sweden[1].

We plot two heatmaps based on the residuals from equally many regressions (see Engström, forthcoming, and GREEN SURGE D4.2 for more details on the method). The heat maps are constructed using an inverse distance type interpolation between the residuals of observations which result from regression analysis[2]. We start with a simple regression analysis featuring only house specific attributes, in order to provide a comparison. The results are depicted in the first map (**Figure 1**), which shows the residuals from a regression analysis of the variables living area, monthly fee, plot area, floor number and building storeys together with time dummies on the contract price[3]. This resulted in an R2 value of 0.61 implying that we are already explaining around 61% of the variation in final prices. The interpretation of the resulting residuals is the deviation from the average price after controlling for all housing specific attributes. Alternatively we can also view these residuals as a representation of neighbourhood values i.e. all neighbourhood specific characteristics affecting final sale prices. The colours of the heatmap range from blue to red indicating whether properties in this area typically sell at a lower value than average (blue) or higher than average value (red). From the map we can thus see that the most attractive neighbourhoods are located in the vicinity of the central station and the waterfront. From this simple regression it is so far hard to see any correlation between the parks and property values.

[1] See GREEN SURGE MS29 and D4.2.

[2] The interpolation analysis was done using the spatial analysis software QGIS.

[3] The regression was running using robust standard errors. Also, the map includes only residuals within two standard deviations we thus avoid extreme values from distorting the map.

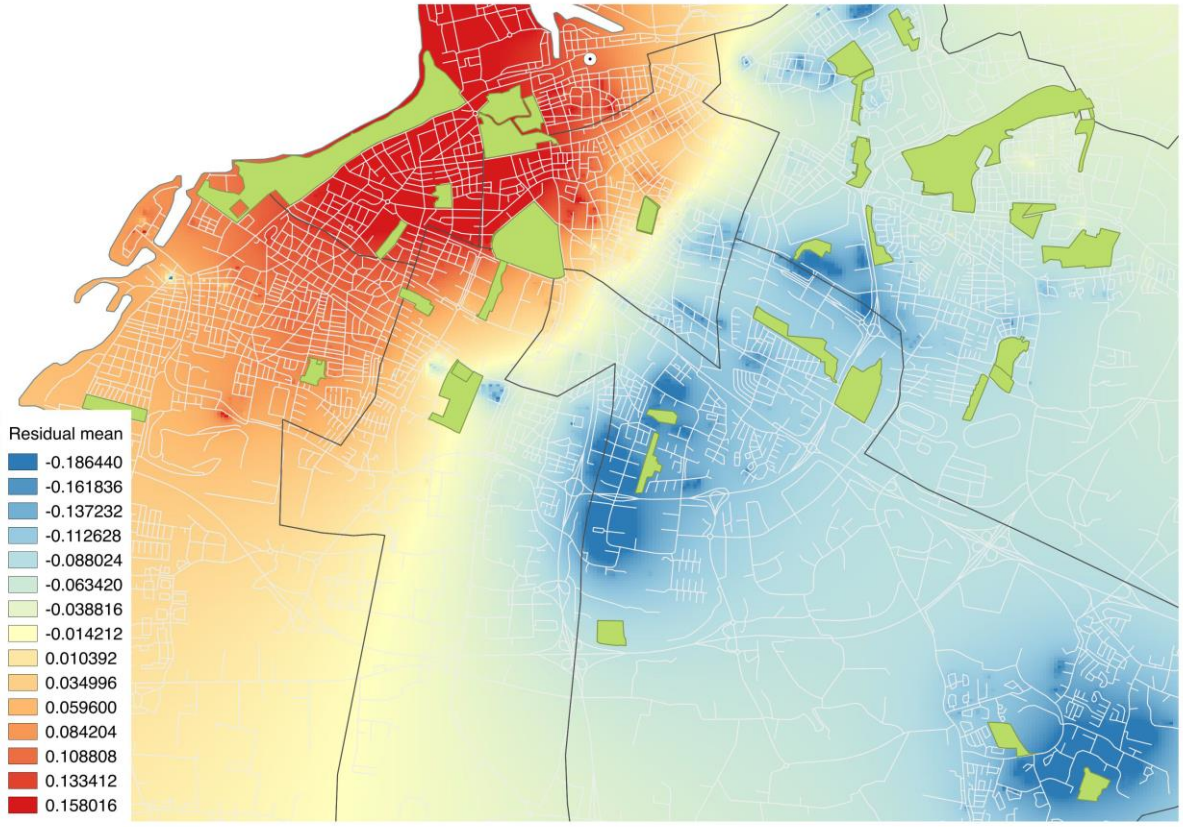

Figure 1. Malmö heatmap based on residuals from regression featuring only housing specific characteristics.

In the second regression analysis we now include neighbourhood specific attributes. By this we mean that we include all variables from table 1 with the exception of park area and park distance. We are thus controlling for several neighbourhood characteristics such as median income, share of non-European population, distance to city centre and waterfront. However, by leaving out the variables park distance and area from our regression, this hopefully reveals more clearly whether the vicinity to a specific park is associated with a higher property value. The residuals from this regression are depicted on the map in **Figure 2**. As can be seen the resulting map shows a lot more colour variation than the previous map. This map thus gives us a better idea of which parks that are more attractive and which are less valued on the housing market. Here we see that in particular the area around the triangular park (Pildammsparken) has a strong red colour, indicating that this is a popular area on the housing market. At the opposite end, east-by-northeast from Pildamsparken, we have the area of Kirseberg where the smaller sized parks do little for the real estate prices in their vicinity[4].

[4] It should however be noted that there may still be a correlation although it may be so weak that it fails to appear on the map.

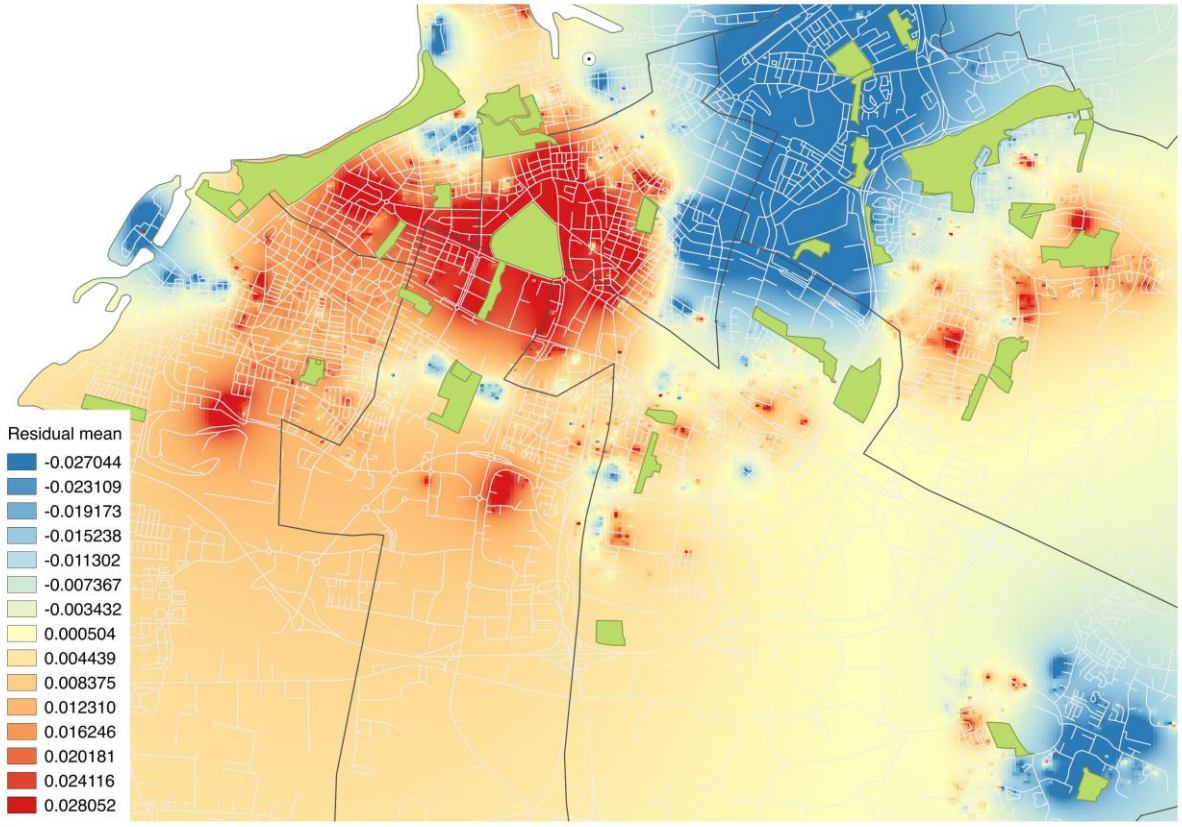

Figure 2. Malmo heat map based on residuals from regression excluding only distance to closest park and park size.

To sum up, the value of heat maps in the way we portrayed them here is that they can tell us something more specific about the spatial distribution of property values as opposed to the more general regression analysis which is the most common method used in the literature.

It should be noted that apart from considerations regarding the functional form for the hedonic function the estimation of a hedonic price function poses a number of other econometric problems. Perhaps the most severe of this problem is the endogeneity problem that arises due to the fact that property values also determine residential development. In other words, unobserved variables that determine real estate prices will likely be correlated with our variable of interest (i.e. the distance to parks). This gives rise to what is commonly referred to as omitted variable bias which as the name suggests biases our estimates of interest. For a description of how to deal with this bias see Engström (forthcoming).

3 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