Do Green Buses Make Property Prices Rise?

A Regression Discontinuity Approach

Olle Norén
23344

Tina Dalaei
23345

Bachelor Thesis in Finance
Stockholm School of Economics
May 15, 2017

Abstract:
Residents in Stockholm had called for an improved physical environment regarding noise reduction as well as better air quality and in August 2014, a fleet of environment-friendly city buses was introduced on a limited number of bus lines in downtown Stockholm. This paper means to examine a possible value created and shown in rising prices of property in the vicinity to the new buses. The study also aims to highlight and quantify the corporate social responsibility, CSR, efforts made by the City of Stockholm as well as those made by the public transportation sector. By using a highly explanatory hedonic pricing model, the authors control for the characteristics valued when people purchase a home. In addition, the remaining variance in the residuals is analyzed using a regression discontinuity (RD) design. In the global RD approach, a rise in property prices by approximately 6% is found. However, the local RD method does not support this, indicating little evidence of increased property value as a result of closeness to environment-friendly buses. It seems that despite the importance of a healthy environment, other factors than the fuel used in the nearby buses, set prices for inner-city apartments in Stockholm.

Keywords: regression discontinuity, hedonic pricing model, public transportation, CSR

Acknowledgements:
We would like to express our gratitude for the valuable input from our tutor Cristian Huse, throughout the development of the thesis. Furthermore, we would like to thank Nikita Koptyug for introducing us to web scraping and Karl Orton for providing us with valuable information about the new buses.

Table of Content

1 Introduction

It is common knowledge that property prices are higher close to public transportation infrastructures. What is less known, however, is whether nearby climate friendly transportation systems has an impact on property prices. In this thesis, we examine the possible value created by the introduction of more eco-friendly buses, commonly called "Green buses". Our aim is to highlight and quantify the *CSR*[1] efforts made by the City of Stockholm and the public transportation sector. Furthermore, we study property prices as an approximation of residents' valuation of improved environmental quality of their surroundings.

The central parts of Stockholm have problems with air and noise pollution. In the entire Stockholm area, there are almost 300 000 bus riderships a day (AB Storstockholms Lokaltrafik, 2016). Furthermore, the number of bus riderships is expected to rise as the population of the city grows. This will result in more traffic and will increase the demand for more environmentally efficient solutions in public transportation. The resemblance between Stockholm and numerous other European cities makes Stockholm an interesting city to study. It has a well-developed public transit system, the population is fast growing and the awareness of environmental issues has increased considerably in the past years. Hence, the results of this paper are applicable to many similar cities, and therefore of value to others.

In August 2014, a fleet of environmental-friendly city buses was introduced on a limited number of bus lines in downtown Stockholm. Following complaints from inner-city residents and measures to reach climate objectives, the aim of the new fleet was to improve air quality and reduce noise pollution. Noise measurements are made every fifth year in Stockholm (Länsstyrelsen Stockholm, 2017), which makes it impossible to quantify marginal sound improvements due to certain actions. Accordingly, we use property prices as an approximation of residents' valuation of improved quality of their surroundings.

A hedonic pricing model consisting of data on 1888 apartments sold between 2012 and 2017 is used to examine the inhabitants' valuation of property characteristics and external factors. Furthermore, a regression discontinuity (RD) design approach is implemented to capture the potential increase in value and is illustrated graphically. The brief conclusion is that there is a price increase of approximately 6%, however, we cannot reliably

[1] Corporate Social Responsibility

prove that the increase is a result of better air quality and less noise caused by the change from ethanol buses to biodiesel-hybrid buses.

A quantification of the inhabitants' appreciation of a cleaner environment is valuable for a number of reasons. Firstly, policy makers can be influenced if they see that this technology also brings monetary value, besides being good for the environment. Secondly, the information can be useful for real estate developers and city planners when deciding on proximity to bus stations.

Previous literature primarily focus on rail ways and the opening of new stations or systems, and hence not on *buses* or *the change of fuel*. Also, in the Scandinavian countries, there is hardly any literature on the topic of public transportation systems and environmental effects. Thus, this study builds on international literature and hopefully it might contribute to further knowledge within this field.

The remainder of this thesis is as follows. Previous research about the effect on house prices, due to public transportation developments, is presented in Section 2. Section 3 provides information of health effects from pollution, and a background of the technology of the buses before and after the change. Data collection and variables are described in Section 4. Section 5 contains a brief review of the used empirical methods, namely hedonic pricing model and RD Design, followed by a hypothesis development. Results are illustrated by tables and graphs, and interpreted in Section 6. The most important conclusions and suggestions for further research are presented in Section 7.

2 Previous Literature

There have been rigorous studies in what determines house prices. The reason for this is out of interest in what people value when purchasing a home and perhaps also in hope to discover undervalued real estate assets. Since properties are products with different attributes and priced accordingly, they are considered to be heterogeneous commodities (McLeod, 1984). The value of these commodities is, from an analytic perspective, a composition of location, rents and extrinsic effects (Krantz, et al., 1982; Hickman, et al., 1984; Shefer, 1986; Strange, 1992; Can, 1992; Dubin, 1998).

One extrinsic effect is access to public transportation. It is well documented that property prices are higher close to public transportation infrastructures (including Bajic, 1983; Baum-Snow & Kahn, 2000; Bowes & Ihlanfeldt, 2001; Damm, et al., 1980; Dewees, 1976; McDonald & Osuji, 1995; Voith, 1993). However, the influence of these kinds of external accessibilities differs among communities, based on what urban amenity is taken into account. For instance, Thériault, et al. (2007) find that households in Quebec City value schools higher than labor markets.

Some studies particularly highlight the effects of air and sound quality. Chen & Whalley (2012) show that the opening of a new urban rail transit in Taipei reduced one key tailpipe pollutant (carbon monoxide) by between 5% and 15%. In another empirical study of the negative effects of noise pollution, Wilhelmsson (2000) finds that single-family houses in Stockholm located near noisy roads sell for a 30% discount. Davis (2008) examines the effect of a policy program in Mexico City which bans drivers from using their vehicles one weekday a week on the basis of the digits on their license plate. Using high frequency measures, he finds that there is no evidence of improved air quality, on the contrary, it resulted in more high-emissions vehicles being in movement.

The effect of *improved* transportation systems, including both railways and buses, has been quantified and evaluated as well. Des Rosiers, et al. (2010) find that an increase of regular bus frequencies results in a value decrease for properties close to regular routes. However, the opposite was true regarding the Express lines. By offering fewer stops and more comfort Express lines proved to be a good alternative to private cars, and thus, prices were influenced positively. Mulley & Tsai (2017) investigate whether a rapid bus transit system, implemented in Sydney in 2003, affects land values or not. Although new transport infrastructure is expected to increase land value through improved accessibility for local residents, this study proves little price difference between properties close to the bus rapid transit stations and those farther from the stations.

The hedonic modeling approach has been widely employed in terms of methodology for capturing the increased value due to a change in transportation systems. This model incorporates property attributes and neighborhood characteristics which also are considered to have an impact on the value of the property (Cervero & Kang, 2011; Concas, 2013; Dube, et al., 2011; McMillen & McDonald, 2004; Mikelbank, 2004; Rodriguez & Mojica, 2009).

To measure the effect of a policy implementation or an intervention, different statistical methods can be used. A common method is *difference in difference*, where the price change due to a new intervention is compared to a control group that did not go through the change. However, the heterogeneity of apartments makes finding a suitable control group extremely hard.

Another way to capture a rise in property value is the repeat sales approach. In this method, the sample contains apartments that have been sold more than once during the time span. The intuition for this is that it controls for all other aspects that determine house prices since they are assumed to be unchanged. The dependent variable then is the difference between the price of an apartment before and after the intervention (Mulley & Tsai, 2017). However, this requires an extremely large data set which makes it difficult to use. Taking these drawbacks into consideration, discontinuity approach, is a suitable method. This method is used by Chen & Whalley (2012) and Bento, et al. (2011), and is further discussed in Section 5.

3 Health Effects from Pollution and Environmental Impact of the New Buses

Policy makers should account for both benefits and costs of polluting activities, nevertheless it can be hard to identify those who benefit and those who are harmed by these activities. The negative byproducts of economic activity are an unavoidable fact which Clay, et al. (2016) show when they investigate the effect of coal power plant emissions on infant mortality in the US during the mid 20th century. They find that there were substantial negative health effects associated with an expansion of the power grid line. The benefits of access to electricity however presented a clear tradeoff against the health costs of unregulated emissions. Given limitations in transmitting the electricity across big distances, many of the households that benefited from the power grid lines were the same to be exposed to the emissions. Access to roads and public transportation face a similar tradeoff, where the benefits of accessible transportation needs to be weighted against the costs of it.

Economic cost benefit analysts of policies and development projects often involve assigning economic values to the changes in the risk of fatality. A common term used in this context is the value of statistical life (VSL). VSL reflects individuals' willingness to pay for risk reduction of fatality and therefore the economic value to society to reduce the statistical incidence of premature loss of human life in the population by one (Viscusi & Aldy, 2003).

In addition to studying Clay, et al. (2016) we have looked into Greenstone & Chay (2003) who investigate the effect total suspended particulates (TSPs) on infant deaths in Mexico City. They find that a 10% reduction of TSPs reduce the number of infant deaths by 800 annually. To be able to connect monetary calculations to environmental efforts, human lives have been given a price which differs between countries (Ashenfelter, 2006). This might by many be seen as immoral and distasteful, nevertheless it is used as an economic measurement. Davis (2008) studies Greenstone & Chay (2003) and in his own report he adds the value of a statistical life. To calculate the infant mortality rate, he uses the birth and infant mortality rate published in the 2006 edition of the World Bank's World Development indicators. According to this, the reduction of infant mortality induced by a 10% TSP improvement alone, would imply benefits of $1.48 billion.

In our study, we have no mortality rates, and thus no figures on the economic benefits of saving lives thanks to an improved environment. However, table 1 illustrates health and environmental damage caused by different pollutants.

Table 1

Substance	Effect on health	Effect on environment
Particles	Impact respiratory diseases, reduction in lung function, worsening of asthma and other lung diseases. Can also contribute to the development of asthma. Increased risk of death by heart and lung diseases and cancer. (Stockholms stad, 2014)	Storage of heavy metals and organic environmental toxin in ground and sediment. Pollution and climate effects.
VOC[2]	Benzene can cause cancer, mainly leukemia. Aldehydes irritates respiratory tracts and may aggravate asthma. (Länsstyrelsen Blekinge Län, 2005)	Formation of ground-level ozone, which damages plants and material. (Länsstyrelsen Blekinge Län, 2005)
NOx	Impacts respiratory conditions, worsens lung function, increases allergens response. (Icopal, 2017)	Damages vegetation. Can also damage vegetation through reaction with other pollutants in the creation of ozone. (Icopal, 2017)
Noise	Temporary negative impact in terms of higher heart rate and increased blood pressure. Permanently, it can also damage hearing ability, cause sleeping problems, stress and worsened concentration and learning ability. (Naturvårdsverket, 2017)	

Note: This table contains health and environmental damage caused by different pollutants.

In mid 2014, the company in charge of the bus traffic in central Stockholm, Keolis, invested in 52 new buses, included in the Hornsberg depot, to primarily operate Kungsholmen and Vasastan, and partly Södermalm. The new biodiesel-hybrid buses, which also were the first hybrid buses in Stockholm (Maasing, 2013), replaced the older ethanol driven ones. The aim of the purchase was to improve the ambient environment around the bus stations by reducing noise and air pollution. The new buses officially started to run August 18, 2014 (Orton, 2017), and fulfill the European Union's highest environmental classification Euro 6.

[2] Volatile organic compounds or VOCs are a large group of organic chemicals that have a high vapor pressure at ordinary room temperature.

Economic cost benefit analysts could also be expected to support investments in other clean technologies such as electric buses since it includes a profitability incitement for companies. However, the electric bus line 73 in Stockholm, between Karolinska Institutet and Ropsten, which originally was a EU project, was considered too expensive to continue even though electric buses are markedly better from an environmental standpoint.

Biodiesel is a fuel adopted for diesel engines. It is created by chemical processing of vegetable and animal oils and fats, converting to fatty acid methyl esters (FAME) or hydrogenated vegetable oil (HVO). These two types of biodiesel have slightly different characteristics regarding environmental and health effects, with advantage for the HVO alternative. However, the new buses use rape methyl ester (RME) (a type of FAME) (Böhlin, 2015), due to the lower costs compared to HVO (Trafikförvaltningen Stockholms läns landsting, 2014).

Old ethanol buses versus biodiesel hybrid or electric buses make a difference for Swedish public transportation companies in more than just the environmental aspect. Besides bad reputation and customer losses, the companies bear the risk of getting legally punished if they do not follow certain regulations. Both air and noise pollution are considered as environmentally harmful activities if they cause disorders that according to medical or hygienic assessment may affect health and which are not minor or purely temporary (9th Chapter 1st paragraph Miljöbalken (1988:808)). The penalty for causing this can be jail or a fine (Chapter 29th and 32nd Miljöbalken). There are also provisions regarding the procurement of vehicles. This is meant to promote and stimulate the market for clean and energy efficient vehicles, and improve the transport sector's contribution to the European Union common environmental, climate and energy policy (Lag (2011:846) om miljökrav vid upphandling av bilar och vissa kollektivtrafiktjänser, compare with Directive 2009/33/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of clean and energy-efficient road transport vehicles). Table 2 illustrates a comparison between the old and new buses in terms of emission costs, using a report made by Ecotraffic (2015), ordered by the Swedish Transportation Administration (Trafikverket). It can be noticed that the emission of carbon dioxide (CO_2) actually is higher for the new biodiesel-hybrid buses than the older ethanol driven ones. Nevertheless, the large improvement is captured by the lower amount of particles, VOC, nitrogen oxides (NO_x) and noise, here measured in SEK. The biodiesel-hybrid buses decrease the cost of these pollutants with approximately 20%.

Table 2

Consumption and emission costs	**Old Buses: Ethanol**	**New Buses: Biodiesel Hybrid (RME)**
Fuel consumption in central Stockholm **(liters/km)**	0.70	0.40
CO_2 emission (**kg/km**)	0.41	0.51
CO_2 cost (**kr/km**)	0.60	0.74
Costs		
Particles **(öre/km)**	7.20	4.50
VOC **(öre/km)**	0.50	0.00
NO_x **(öre/km)**	1.80	2.30
Noise **(kr/km)**	4.15	3.30
Total cost	**4.245**	**3.368**

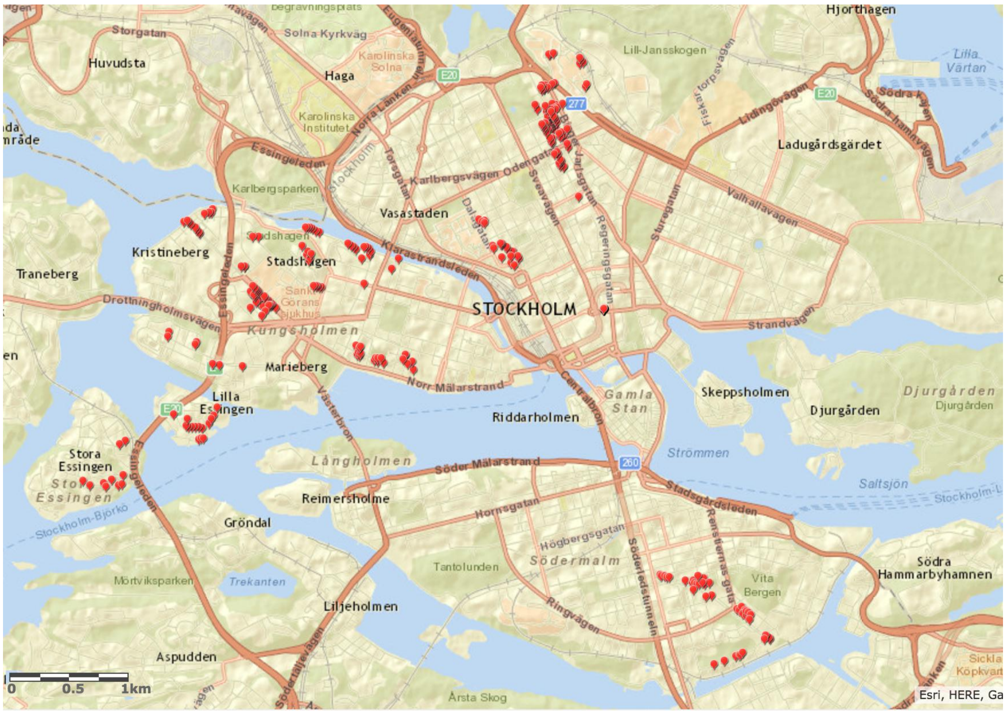
Note: This table illustrates a comparison between the old and new buses in terms of emission costs

4 Data

In this section, all variables which we originally considered using in the hedonic pricing model and in the RD Design, are presented and explained. This section also contains information regarding the data collection process.

The complete data set consists of 1888 apartments that were sold between the years 2012 and 2017 located on 305 different addresses in central Stockholm. Figure 1 plots the addresses on a map showing central Stockholm. The apartments in the data set were selected if they were either directly or indirectly located on streets trafficked by any of the five bus lines that changed to cleaner fuel. To isolate the effect, the data set does not contain any streets trafficked by buses that did not go through the transition to cleaner fuel. The term *directly located* implies that the apartment is facing the street where the bus passes and is within close distance to the bus stop[3]. *In-directly located* implies that the apartment is facing the same street but is further away from the bus stop, or facing a street which is connected to the street where the bus passes. We have made this distinction in belief that apartments closer to where the bus stops are more affected by its noise and pollution.

Figure 1



A map of the locations of the observations in our dataset[4]

[3] We define close distance as approximately 50 m.
[4] A more detailed map of the observations can be found in A1 in Appendix.

The data has been cleaned from outliers in the dependent variable, price. This is to prevent extravagant apartments, where prices are determined by very specific features, to affect the estimated model. These features could for instance be a large roof terrace, indoor pool etc.

Data on sold apartments is available on broker websites such as Booli.se and Hemnet.se, for people to check prices on sold apartments in the area where they consider purchasing a home. The data is available, free of charge, to scroll through and for personal use. However, it is not available for downloading. The easiest method to collect this data would simply be to copy and paste from the web browser. Unfortunately, the amount of data required to do regression analysis would be extremely time consuming to collect this way. A more efficient way to extract this kind of publicly available data is to use a method called *web scraping*.

Web scraping is a software technique for extracting information from websites. This technique focuses on the transformation of unstructured HTML data on web pages, into structured data that can be stored in a local database and analyzed in statistical software packages. Web search and extraction of information is typically performed with a method called *web crawling*. A web crawler is a program or automated script that browses the web at a systematic automated manner. Web scraping is a more recent variant of web crawling, where the aim is to look for specific information, such as prices of particular goods from online retailers, or as in our case, information on sold apartments in Stockholm (Vargiu & Urru, 2013).

To conduct the web scraping we used a programming software called *c#* to produce the automated script for data extraction (Monson, 2009).[5] The software then extracted the data from Booli and exported it into a spreadsheet for further use.

The data extracted from Booli includes the price of the apartment, living area, monthly fee, number of rooms and what floor the apartment is located on. This data is the basis for the hedonic model, and is described in more detail in Table 3. Furthermore, the data from Booli has been completed with additional variables that are commonly known for affecting the final price of an apartment. These variables must to a certain extent been added manually, as well as by using software such as Microsoft Excel add-ins DataStream and Visual Basic.

---

[5] The code we produced can be found in A11 in Appendix.

Table 3

Variable	Obs	Mean	Std. Dev.	Min	Max
Price (SEK)	1 888	3,900,041	1,730,850	1,400,000	12,400,000
Space (m^2)	1 887	52.1	22.3	18	141
Fee (SEK)	1 888	2609.2	1088.0	307	7726
Floor	1 617	3.0	1.8	1	10
Rooms	1 888	2.0	0,85	1	5

Note: Descriptive statistics on the data set which is the basis for the hedonic model

Space – The living area is one of the top features of an apartment. In our hedonic model, we have included the variable space as well as a squared variable, space^2. This is to capture the non-linear effect the living area has on prices (i.e the price difference between a 20 m^2 and 30 m^2 apartment is bigger than a 90 m^2 and 100 m^2 apartment).

Rooms – The number of rooms in an apartment is included since it dictates how many people can live in the apartment. Even if an apartment is spacious, it is difficult to share with others if it does not have rooms for each resident. Unfortunately, some listings on Booli lack data on this variable, resulting in missing values.

Fee – The variable fee is the monthly fee the owner of the apartment has to pay to the tenant owner association. The tenant owner association owns the building in which the apartments are situated, and is managed by its members, the residents. The monthly fee is mainly used to pay interest and amortization on loans, but also maintenance costs. In general, the fee can be considered as a reflection of the financial situation in the association. A high fee can indicate a weak financial situation (high indebtedness), and thereby lower the price of the apartment. We also test the value of fee^2, for the same reasons (non-linearity) as the variable space^2.

Floor – What floor the apartment is located on has an impact on the price. Lower level apartments tend to be cheaper because they are easier to see into and they are more affected by street noise. Higher level apartments can be unattractive as well since many apartment buildings in Stockholm are old and lack elevators.

Subway proximity – The fastest way to travel in Stockholm is using the subway which makes proximity to the closest subway station a natural component of the model. This variable is calculated as the walking distance from the apartment to the nearest subway station. There are several papers researching the effect on housing prices after an introduction of nearby high speed rail communication (e.g. Chin & Chau, 2003). The data on distances is manually collected using a geocoding application in Microsoft Excel Visual Basic to make use of Google Maps distance services.

Location – This distance is measured as walking distance from the Central Station in Stockholm. It is collected the same way as the Subway Proximity and we use it as a dummy variable. The dummy variable equals one (1) if the observation is more than 2 km from central Stockholm. Central Stockholm is in this context the Central Station. A publication from Booli shows the decline in housing prices as the distance from central Stockholm increases (Booli Search Technologies AB, 2016).

The two-year mortgage rate – We collect data on the two-year interest rates on mortgages provided by Danske Bank via DataStream. Interest rates impact the monthly cost of owning an apartment, and could therefore be a valid variable to include in the model. However, this effect could be captured by time fixed effects and/or time trends.

Real estate pricing index (flats in Stockholm) – This index holds the effect of the general price increase of real estate in Stockholm fixed. This variable prevents us from rejecting the null hypothesis when it is merely the prices in general that have risen. It captures both the price increase that is due to a higher demand for apartments as well as the price increase due to inflation. This effect could also be captured by using a polynomial time trend, and/or including seasonal time variables, as well as time fixed effects.

5 Theory and Methodology

In this section, we present our hypothesis development and a review of the theory covering the statistical methods that we use. The hedonic pricing model is a method to determine what factors are important and valued when people purchase an apartment. Furthermore, the regression discontinuity design makes use of the residuals from the hedonic model in order to quantify the new buses' value.

5.1 Hypothesis Development

Hypothesis 1: *The value of less noise and cleaner air is incorporated in real estate prices.*

Our belief/theory is that people value a calmer and cleaner environment in a similar way that they value a beautiful view, a good location and other apartment characteristics. In an efficient market, all aspect that affect the value of an asset should be included in its price (Fama, 1970). Since apartments are sold in a market where the price is determined by supply and demand, the same principles should be applicable. A reduction in noise is intuitively a change that is more noticeable than a change in air quality to the people that live nearby. By distinguishing between directly and indirectly located apartments we investigate whether proximity to the bus stop makes a difference or not. This leads to our second hypothesis:

Hypothesis 2: *Properties located directly by the bus stop are more affected than properties that are not located directly by the bus stop.*

5.2 The Hedonic Pricing Model

The hedonic pricing model is a simple method with an underlying goal to create an accurate model for predicting the price of property.

Buildings and apartments can be compared to a bundle of goods sold in a market. An apartment can be decomposed into characteristics such as number of rooms, what floor it is located on, or distance to the city center. Each of an apartment's characteristics can be seen as items that sum up to the expected overall price. However, unlike a bundle of goods, the value of an apartment's different features cannot be directly observed (e.g. the value of an extra bedroom relative to the overall value of the apartment). The hedonic model can be used to measure the influencing effect of these characteristics on the overall transaction price. They might add or subtract value from the property.

The regression analysis in the hedonic model is multivariate regression using the ordinary least squares (OLS) method. This method minimizes the sum of the squared residuals (the difference between the observed and the predicted prices):

$$min \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{N} (Y_i - \hat{\beta} - \widehat{\beta_i X_i})^2,$$

where Y_i is the observed final price of an apartment, β_i are the estimates and X_i the explanatory variables. The explanatory variables are characteristics of an apartment that we have chosen if

i. We believe they have explanatory power when trying to value a property, and
ii. there is data available for them.

In the process of choosing the hedonic model that explains the prices of the properties in our data set, various combinations of different variables are considered. The four most suitable combinations we find, are built as follows: Y_{it} is the price of the property, 1(BDH) is an indicator variable for properties sold after the change[6], γ_{it} is the time fixed effect in terms of zip code over year and half year, and ε_t is the standard error term which is clustered by zip code times year and week.

The four equations in question are:

$$Y_t = \beta_0 + \beta_1(BDH_t) + \beta_2 X_{it} + \gamma_{it} + \varepsilon_t \quad (1)$$

where X_{it} in equation (1) is the vector of covariates of the variables *space*, *fee*, *rooms*, *walking distance to the closest metro* and *to Stockholm Central Station (integer variables*[7]*)*, and *floor*.

$$Y_t = \beta_0 + \beta_1(BDH_t) + \beta_2 X_{it} + \gamma_{it} + \varepsilon_t \quad (2)$$

[6] 1(BDH) is a dummy variable, which takes the value 1 if the apartment is sold after the change and the value zero otherwise.
[7] An integer variable creates a dummy variable for each value of the variable.

where X_{it} in equation (2) is the vector of covariates of the variables *space*, *fee*, *rooms*, *walking distance to the closest metro* and *to Stockholm Central Station* (integer variables), *floor* (integer variable), and space^2,

$$Y_t = \beta_0 + \beta_1(BDH_t) + \beta_2 X_{it} + \gamma_{it} + \varepsilon_t \tag{3}$$

where X_{it} in equation (3) is the vector of covariates of the variables *space*, *fee*, *walking distance to the closest metro* and *to Stockholm Central Station* (integer variables), *space^2*, *floor* (integer variable), *rooms* (integer variable), *fee^2*, and *quarter of year* (integer variable),

$$Y_t = \beta_0 + \beta_1(BDH_t) + \beta_2 X_{it} + \gamma_{it} + \varepsilon_t \tag{4}$$

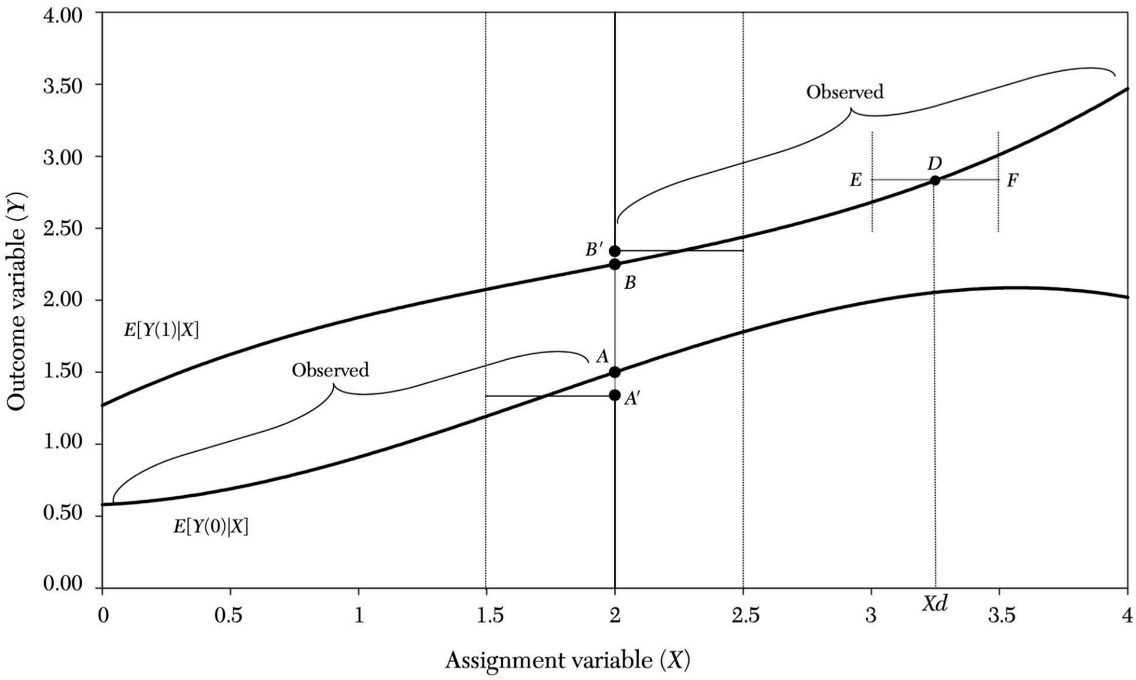
where X_{it} in equation (4) is the vector of covariates of the variables *space*, *fee*, *space^2*, *floor* (integer variable), *walking distance to the closest metro* and *to Stockholm Central Station* (integer variables), *rooms* (integer variable), and *quarter of year* (integer variable).

The coefficient of special interest is β_1, which seeks to give us an estimate of the uplift in properties due to the change of buses. Since we through the hedonic model want to find a model that is appropriate over several years, it is important to include time-fixed effects.

5.3 Regression Discontinuity Design

Thistlethwaite & Campbell (1960) first introduced the regression discontinuity designs (RDD) in order to estimate and analyze how future academic outcomes is affected by merit awards. These awards were allocated based on observed test scores: those that were above a known cutoff point received the awards. The core idea of this research design was that students with scores just below and just above the cutoff should be good comparisons (Lee & Lemieux, 2010). Thus, all other components were assumed to be continuous with respect to the assignment variable X. Since one cannot simultaneously observe the same unit (for instance, student) both before and after the treatment, the average treatment effects over populations are in focus, rather than those on unit level (Lee & Lemieux, 2010).

Figure 2



The treatment effect could be illustrated by the distance between B and A, shown in Figure 2. The quantity, "the average treatment effect", could be estimated by

$$B - A = lim_{\varepsilon \downarrow 0} E[Y|X = c + \varepsilon] - lim_{\varepsilon \uparrow 0} E[Y|X = c + \varepsilon],$$

which equals

$$E[Y_t(1) - Y_t(0)|X = c].$$

It is implemented through a dummy variable taking the value 1 if the unit is on the right side of the cut-off, and otherwise 0:

$$W_i = \begin{cases} 1 & X_i \geq c \\ 0 & otherwise \end{cases}$$

Above equation illustrates the so called *sharp RD Design*, where the units receive the treatment just by passing the cut-off: the probability of receiving the treatment goes from zero to one. In this thesis, a variant of RD Design, *the fuzzy RD Design*, is applied in order to capture that it is the probability that some unit is treated that changes discontinuously when it passes the cut-off, where the probability could increase from 0 to less than 1. This is because we do not expect the market to react immediately on the change, but with a delay, which would not affect all properties simultaneously.

The impact of the treatment effect in the fuzzy RD Design is estimated through the change in outcome either side of the cut-off divided by the change in treatment either side of the cut-off,

$$\tau_F = \frac{lim_{\varepsilon \downarrow 0} E[Y|X = c + \varepsilon] - lim_{\varepsilon \uparrow 0} E[Y|X = c + \varepsilon]}{lim_{\varepsilon \downarrow 0} E[Y|D = c + \varepsilon] - lim_{\varepsilon \uparrow 0} E[D|X = c + \varepsilon]}$$

where c is the cut-off, Y is the outcome variable, X, is the assignment variable and D is the treatment dummy variable, taking the value zero for units being on the left side of the cut-off, and one for units on the right side.

5.3.1 Graphical Analysis

Before running regressions, users of RD Design want to illustrate the raw data and treatment effect graphically in an intuitive way. This is since if the discontinuity is not visible, there will presumably not be any effect captured by the more advanced statistics that follow (Brewer, 2015). Other advantages of the RD graph are that the magnitude of the treatment effect (the "jump") can be illustrated, and it also shows if any unexpected jumps occur at other points in time.

The RD graphs illustrate the behavior of the estimated regression function, where $X_i < c$ being the control group and $X_i \geq c$ being the treatment group. The construction of the graphs is usually made by "dots" being local sample means over non-overlapping bins, together with estimated global polynomial regression curves on both sides of the cut-off.

This thesis includes a graphical illustration of different bins used when the raw data is plotted. Our benchmark for the size of the bins is 7 days, which corresponds to 100 and 133 bins on the right and left side of the cut-off, respectively. However, the estimates will be imprecise if the size of the bins is too narrow (when the number of bins too large). On the other hand, if the size of the bins is too wide (when number of bins is too small), the estimates may fail to estimate the regression slope since they are biased (Lee & Lemieux, 2010). Therefore, we also illustrate the graphs of the doubled and halved the size of the bins twice, where the size of the bins approximately is 29 and 14 days, as well as 3.5 and 2 days.

5.3.2 Formal Regression

The formal regression is made following two different approaches: global polynomial regression and local linear regression.

5.3.2.1 Global Polynomial

The parametric *global polynomial* approach within the RD Design, is simply an OLS estimation, where the coefficient of the treatment-dummy variable is the coefficient of interest. The regression model used in this thesis is the best hedonic model found by considering different variables affecting property prices. After choosing which variables to include in the hedonic model, the model is tested for different time trends, with the aim to capture underlying time variation of the data set and increase the explanatory power of the hedonic regression model.

The hedonic regression model, including the time trend variable, is run, and the *residuals* are saved. By considering the residuals of the property prices rather than the prices themselves, the unexplained portion of the property prices is captured, which intuitively should include the value of environmental qualities, thereby the buses' switch to cleaner fuel. Thus, it is the residuals that are plotted in the RD graphs that follow.

When performing global polynomial RD Design, there might be a concern of an unbalanced number of observations on each side of the cut-off (Bento, et al., 2011). In the following section, we will present a local linear methodology with intention to confirm the global polynomial results.

For further knowledge and reading about global polynomial estimations within RD Design, papers by Davis (2008) and Bento, et al. (2011) are recommended.

5.3.2.2 Local Linear

Regarding the non-parametric approach, *local linear/polynomial regression*, our estimations are made in two steps. First, the hedonic pricing model - including the same appropriate time trend as in the global polynomial - is performed across the whole sample and the residuals are saved. Then, we use the residuals to perform the RD analysis. The local linear approach views the treatment effect as a local randomization and the analysis is limited to observations close to the cut-off point, located within the so-called *bandwidth* (i.e. “window”). Since the local analysis only uses observations near the cut-off point, the sample size is smaller in relation to the global polynomial analysis. Thus, the functional form of this regression is expected to be close to linear. In this thesis, the bandwidth is chosen after visual examination of the distribution of the assignment variable day, since we do not have a natural choice of window around zero. Another alternative would be to adopt data-driven bandwidth selection methods, which seek to minimize the standard errors, taking into account whether the design is sharp or fuzzy (see for instance Calonico et al., 2017).

Choosing the appropriate bandwidth is a trade-off between precision and bias. Larger bandwidths yield more precise estimates since it includes more observations in the regression. However, it is less likely that the functional form will be linear, and thus, it could bias the estimate of the treatment effect. In this thesis, different bandwidths are considered and tested for. The shortest time it would take for inhabitants to react to less noise and cleaner air through improved buses is assumed to be two weeks (15 days), why the smallest tested bandwidth is 15 days on each side of the cut-off. The widest bandwidth tested for is 70 days on each side of the cut-off. A data-driven bandwidth selection generates bandwidths between 180 and 400 days on each side of the cut-off, which in our case is assumed to be too wide. The kernel function used in order to estimate the local polynomial is triangular, which is the default one. In practice, the choice of kernel function is expected to have little influence (Lee & Lemieux, 2010).

For further knowledge and reading about local linear regression within RD Design, papers by Lee & Lemieux (2010), Calonico, et al. (2014) and Calonico, et al. (2017) are recommended.

6 Results

In this section, we present our main findings. The best hedonic pricing model is chosen and explained. The outputs of the RD Design are presented for the whole sample as well as for a division between properties located directly and indirectly by the bus stops. The global approach provides some evidence of the new buses being priced by the market, still, the local approach does not support the global findings. Lastly, we perform sensitivity and robustness checks concerning the RD Design.

6.1 Hedonic Pricing Model

Table 4

Equation	(1)	(2)	(3)	(4)
VARIABLES	price	price	price	price
1.(BDH)	253,339**	230,676**	175,910**	175,559**
	(118,238)	(116,239)	(86,666)	(86,546)
			(149,763)	(149,650)
Constant	78,035	140,345	732,303***	729,792***
	(135,170)	(132,637)	(160,789)	(152,834)
Observations	1,617	1,617	1,617	1,617
R-squared	0.945	0.945	0.949	0.949
Adj. R-Squared	0.930	0.930	0.935	0.935

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This is a reduced version of table A2 which can be found in the Appendix. It reports the estimates from four separate regressions of equations (1), (2), (3) and (4), that are made in order to capture the best fitting hedonic pricing model for the data set. The dependent variable is the property price in SEK (collected April 1st, 2017). The coefficients of the regression variables should be interpreted in monetary terms, SEK. Standard errors are in parentheses and P-values of the OLS are in square brackets. R-squared ranges from 0.945 to 0.949, and adjusted R-squared ranges from 0.930 to 0.935.

Table 4 shows the output from the four most appropriate regression models. When choosing the best hedonic model to use in the RD Design, the Bayesian Information Criterion (BIC) is used as benchmark. The lowest BIC-value, and consequently the most appropriate model according to the criterion, is received by equation (4). Equation (4) also has a coefficient of 1(BDH) that is within reasonable amounts. We expect a positive coefficient of the treatment variable, as an approximation of residents' valuation of a cleaner local environment. Hence, we choose equation (4) as the best fitting hedonic model.

As illustrated in A2 in Appendix, all variables in equation (4) except number of rooms and floor level are significant at either 1% or 5% significance level, indicating that they have been of high importance for people in search for an apartment to purchase. It is important to be aware of the seasonal trend in housing supply. As A3 in Appendix depicts, the property transactions are not evenly distributed throughout the year. Equation (4) includes an integer variable for each quarter to control for this effect.

A regression of equation (4) generates an adjusted R-squared over 0.93, which indicates that more than 93% of the variation in the dependent variable closing price can be explained by the chosen independent variables. The remaining 7% of the variation is within the residuals, which will be further examined in the RD Design.

The fixed effects that we include in equation (1) - (4) concern zip code, year and half year. This is to control for the unobserved factors within each zip code, factors that are expected to stay constant over time.[8] We test whether all these time fixed effects are jointly equal to zero, receiving an F-value of 157.06, which allows us to reject the null hypotheses that they are simultaneously zero. Including these fixed effects, economically means that we are able to adequately capture unobserved heterogeneity.

[8] A table documenting the number of transactions in each zip code per half year is found in A4 in Appendix.

6.2 The Whole Sample

6.2.1 Global Polynomial

Table 5

Model	(1)	(2)	(3)	(4)	(5)
VARIABLES	price	price	price	price	price
1.(BDH)	157,769*	174,354**	180,597**	183,780**	185,752**
	(88,182)	(86,536)	(86,038)	(85,975)	(85,958)
Constant	488,635*	706,211***	734,615***	739,979***	740,387***
	(276,732)	(155,316)	(153,053)	(154,012)	(153,969)
Polynomial Time Trend	1	2	3	4	5
Observations	1,617	1,617	1,617	1,617	1,617
R-squared	0.950	0.950	0.950	0.950	0.950
Adj. R-squared	0.935	0.935	0.935	0.935	0.935

Model	(6)	(7)	(8)	(9)
VARIABLES	price	price	price	price
1.(BDH)	186,959**	187,603**	187,817**	187,699**
	(85,940)	(85,933)	(85,941)	(85,963)
Constant	738,970***	736,694***	734,029***	731,276***
	(153,593)	(153,211)	(152,923)	(152,746)
Polynomial Time Trend	6	7	8	9
Observations	1,617	1,617	1,617	1,617
R-squared	0.950	0.950	0.950	0.950
Adj. R-squared	0.935	0.935	0.935	0.935

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: This table reports the global estimates from nine separate regressions of equation (4), including each of the first, second, third, fourth, fifth, sixth, seventh, eighth and ninth order of polynomial time trend, respectively. The coefficients of the regression variables should be interpreted in monetary terms, SEK. Standard errors are in parentheses.

Table 5 reports the outcome of nine separate regressions of equation (4) with different polynomial time trends. To ensure that the effect is captured in the residuals, the treatment variable 1(BDH) is excluded from this regression. The coefficient of interest, β_1, is significant at a 5% significance level for all time trends, except for the first order of polynomial, at which it is significant at a 10% significance level. All regressions generate an R-squared and adjusted R-squared of above 0.93. The BIC is used as benchmark when deciding which polynomial of order to choose. However, the BIC tells us that all nine regressions are equally good[9].

In order to find the best fitting time trend, we compare scatterplots of residuals from the regressions reported in Table 5. Figure 3 (a) provides the scatterplot of residuals from a fifth order polynomial time trend, which by visual examination proved to be the most suitable one[10]. The fitted line smoothly follows the plotted residuals, implying that the underlying time trend is adequately captured by the fifth-order polynomial. By studying Figure 3 (a) and (b) one can see that the raw data is relatively smooth, which indicates a fair OLS estimation. Furthermore, we notice a discontinuity in the property prices at the intercept, 18th of August, 2014. Figure 3 (b) illustrates the same residuals as in Figure 3 (a), but includes confidence intervals of 95% coverage for each bin. Once again, the smoothness of the data set can be visualized since the confidence intervals of the residuals are rather tight, implying low variance.

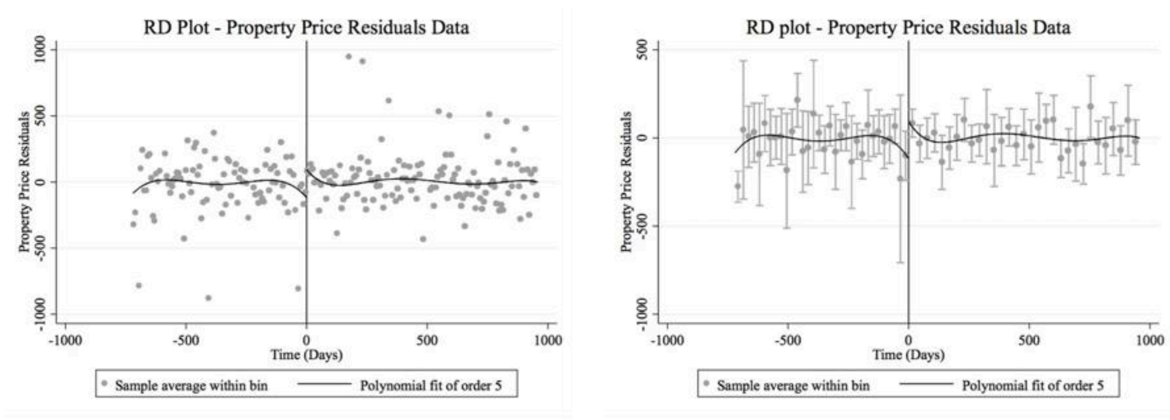
With a fifth order of polynomial trend, the uplift in property prices due to the switch to cleaner fuel is approximately 185,000 SEK[11]. To put this in perspective, the average price of apartments during the two months leading up to the change was 3,048,036 SEK, which implies an increase of 6.1%.

The results of the global polynomial analysis provide evidence on the 5% significance level that the uplift in the property prices is due to less noise and cleaner air. This allows us to reject the null hypothesis that there was no effect brought by the change of buses. Our conclusion is further confirmed by the visible discontinuity in Figure 3.

[9] The difference between the BIC of these nine regressions is less than 2 in absolute terms, which according to (Raftery, 1995) provides low evidence that one is better than another one.
[10] See Appendix A5 for illustration of different orders of polynomial time trends.
[11] This is the value of the coefficient of 1(BDH) with a fifth order of polynomial time trend, seen in Table 5.

Figure 3

(a) (b)



*Note: **Figure 3 (a)** plots the residuals (in thousands SEK) from regression of equation (4), excluding 1(BDH), including a fifth order of polynomial time trend, where each bin is 7 days. See A6 Appendix for a comparison of different bin lengths. **Figure 3 (b)** plots the residuals from regression of equation (4), excluding 1(BDH), including a fifth order of polynomial time trend, illustrating a 95% confidence interval for each bin. The bin sizes are automatically chosen, mimicking the underlying data variability: each bin is 22.5 days on the left side and 30.8 days on the right side of the cut-off.*

6.2.2 Local Linear

Table 6

Bandwidth:	**15**	**30**	**40**	**50**	**60**	**70**
Polynomial of Order:						
One	172,767	176,139	150,207	156,533	215,033**	247,959***
	(174828)	(118489)	(115860)	(109216)	(88073)	(79819)
	[0.323]	[0.137]	[0.195]	[0.152]	[0.0146]	[0.00189]
Two	563,693	88,841	201,912	168,247	107,507	127,982
	(567047)	(208728)	(193541)	(150648)	(142056)	(129964)
	[0.320]	[0.670]	[0.297]	[0.264]	[0.449]	[0.325]
Three	-107,190	189,199	-135,321	185,534	225,077	118,002
	(1,296,000)	(435403)	(325452)	(251650)	(213248)	(161165)
	[0.934]	[0.664]	[0.678]	[0.461]	[0.291]	[0.464]
observations	*24*	*58*	*75*	*88*	*104*	*127*
	*** p<0.01, ** p<0.05, * p<0.1					

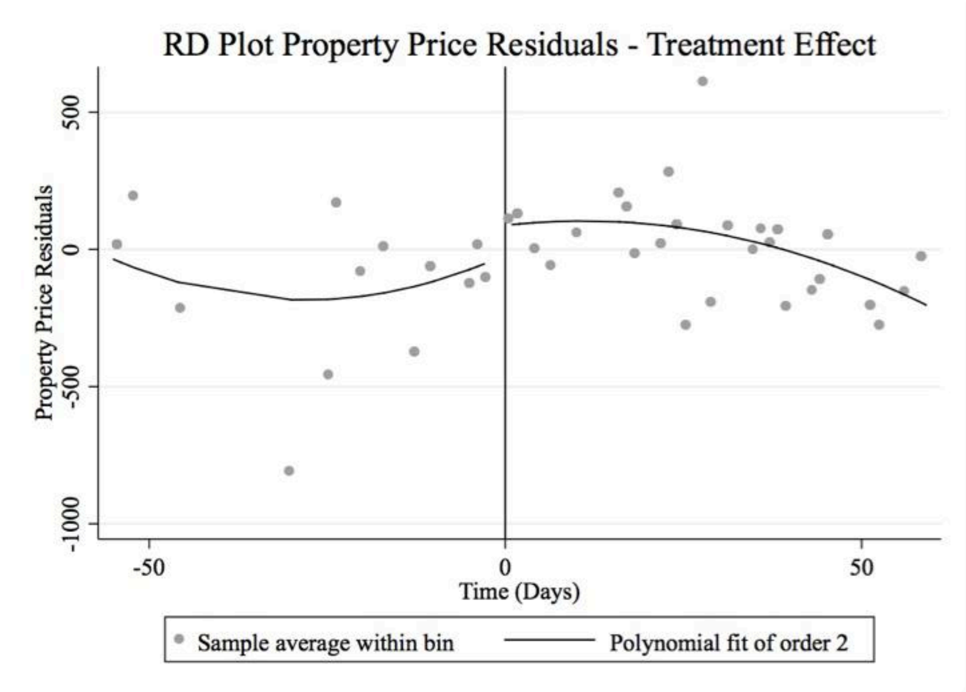
Note: This table reports the local estimates from 24 separate regressions of the residuals in equation (4), excluding 1(BDH), along with a fifth order of polynomial, using RD robust. The order of polynomial is based on the Bandwidths are in number of days. The coefficients of the regression variables should be interpreted in monetary terms, SEK. Conventional standard error in parenthesis, and conventional p-value in brackets.

Table 6 reports the estimates of the treatment effect of the change to cleaner fuels. To ensure completeness, a range of bandwidths and polynomial orders are included. Since we expect the functional form of the local RD approach to be close to linear, we include the estimations from the first (i.e linear), second and third order of polynomial time trend. The estimates indicate that there is generally little evidence of a discontinuity at the intercept. Although the first order of polynomial has statistical significance for bandwidths 60 and 70 days, the explanatory value of the estimates can be questioned. As expected, the precision of the estimates improves as we approach larger bandwidths, and thereby more observations. The significance can therefore be explained by the larger bandwidths. What is more, the size of the 1(BDH) coefficient might seem unrealistically large. We also notice that as the order of polynomial increases, the estimates vary more between different bandwidths.

Figure 4 illustrates the treatment effect by plotting the residuals from equation (4), excluding 1(BDH), along with a fifth order of polynomial. The graph only shows the part 60 days before and after the cut-off, which is why it takes the shape of a second order polynomial. The discontinuities indicated in Figure 4 are consistent with the estimates reported in Table 6 for a second order polynomial with a bandwidth of 60 days. We observe that a second order of polynomial explains the data relatively adequately. Yet, it is not as smooth as illustrated in Figure 3, which is a consequence of using the local method. It is visible that there are fewer observations on the left side of the cut-off, which again can be explained by the seasonal trend in the housing market.

Accordingly, the local RD approach does not reveal enough significant evidence to reject the null hypothesis that the value of less noise and cleaner air is not incorporated in real estate prices.

Figure 4



Note: ***Figure 4*** *plots the residuals (in thousands SEK) from regression of equation (4), excluding 1(BDH), including a fifth order of polynomial time trend, with the size of each bin being one (1) day on each side of the cut-off. Bandwidth is manually chosen to 60 days on each side of the cut-off.*

6.3 Dividing into Directly and In-Directly

6.3.1 Global Polynomial

Table A7 (a) and (b) in the Appendix illustrates global polynomial estimates of regressions made based on the location of the properties. The coefficient for 1(BDH) is insignificant for properties located directly by bus stops, while it is significant for those located farther away from the bus stops, for all orders of polynomial. This is the opposite to what hypothesis two suggested. However, it is likely due to the difference in number of observations, where there are more apartments farther away from bus stops in our data set. The BIC, once again, indicates that all orders of polynomial are equally fit to include in equation (4). Based on this, we pick the best fitting time trend to include for each group of properties by using visual analysis of the graph.

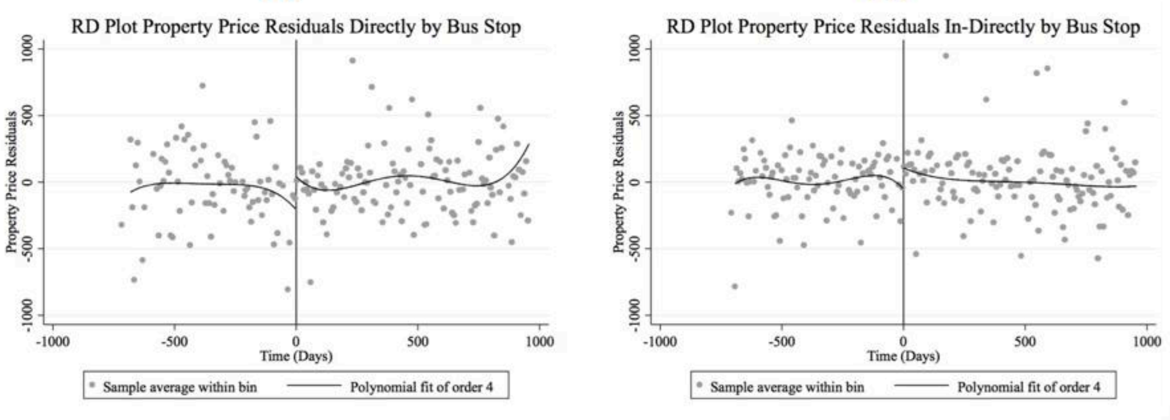
As illustrated in Figure 5, the fourth order of polynomial time trend adequately describes the underlying data, both regarding properties directly located by the bus stops, as well as indirectly.[12] Moreover, the results are smoother for properties located in-directly (Figure 5 (b)) in comparison to those located directly by the bus stops, which supports the findings in Table A7.

Hence, the results of the global polynomial analysis provide no evidence that the uplift of properties directly located by the bus stops is due to less noise and air pollution. However, the results confirm on a 10% and 5% significance level, that the increase on properties located in-directly by bus stops are due to the switch to cleaner fuel. These results are also in line with the illustrations in Figure 5.

[12] See A9 and A10 in Appendix for illustration of different orders of polynomial time trends.

Figure 5

(a) **(b)**



Note: ***Figure 5 (a)*** *plots the residuals from regression of equation (4) for properties located directly by bus stop, excluding 1(BDH), including a fourth order of polynomial time trend, with the size of each bin being 7 days.* ***Figure 5 (b)*** *plots the residuals from regression of equation (4) for properties located in-directly by bus stop, excluding 1(BDH), including a fourth order of polynomial time trend, with the size of each bin being 7 days.*

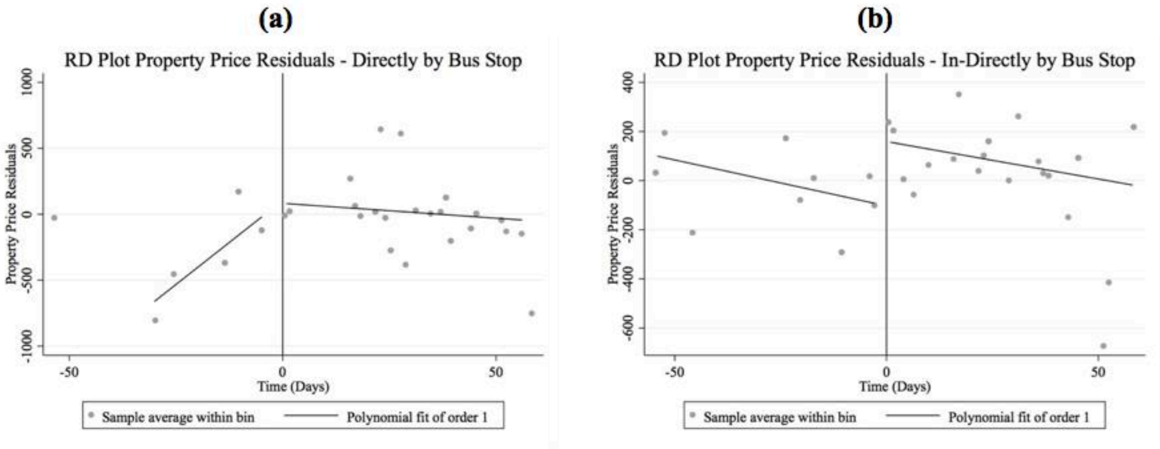
6.3.2 Local Linear

Table A8 in the Appendix reports the estimates of the treatment effect of the change to cleaner fuels, divided into whether the properties are located directly or in-directly relative to the bus stops. As in Section 6.2, a range of bandwidths and polynomial orders are included for completeness. We see that none of the estimates for properties directly by bus stops are significant, implying no evidence of a price increase as a result of the change of buses. Also, as captured by the global method in Section 6.3.1, the local approach provides evidence on 10%, 5% and 1% significance level for the first order of polynomial time trend, indicating a price uplift by more than 250,000 SEK. In addition, we observe that, regarding properties located in-directly by bus stops, the first order of polynomial time trend is best explaining the underlying data distribution. This fulfills the expectation of the local linearity.

Figure 6 illustrates the treatment effect by plotting the residuals from equation (4), excluding 1(BDH), along with a fourth order of polynomial over the whole data set. The figure only shows the part 60 days before and after the cut-off, and due to the low number of observations, it takes the shape of a first order polynomial. A first order of polynomial explains the data relatively adequately. It is not as smooth as Figure 5, which is a consequence of a small bandwidth of 60 days.

The local linear approach provides some evidence of price increases of approximately 250,000 SEK for properties in-directly located by bus stops. No such evidence is found regarding the properties directly located by bus stops. Furthermore, we cannot reject the null hypothesis that properties located directly or in-directly are affected the same way by the change of buses.

Figure 6



*Note: **Figure 6 (a)** plots the residuals (in thousands SEK) from regression of equation (4) for properties located directly by bus stop, excluding 1(BDH), including a fifth order of polynomial time trend, with the size of each bin being one (1) day on each side of the cut-off. Bandwidth is manually chosen to 60 days on each side of the cut-off. **Figure 6 (b)** plots the residuals from regression of equation (4) for properties located in-directly by bus stop, excluding 1(BDH), including a fourth order of polynomial time trend, with the size of each bin being one (1) day on each side of the cut-off. Bandwidth is manually chosen to 60 days on each side of the cut-off.*

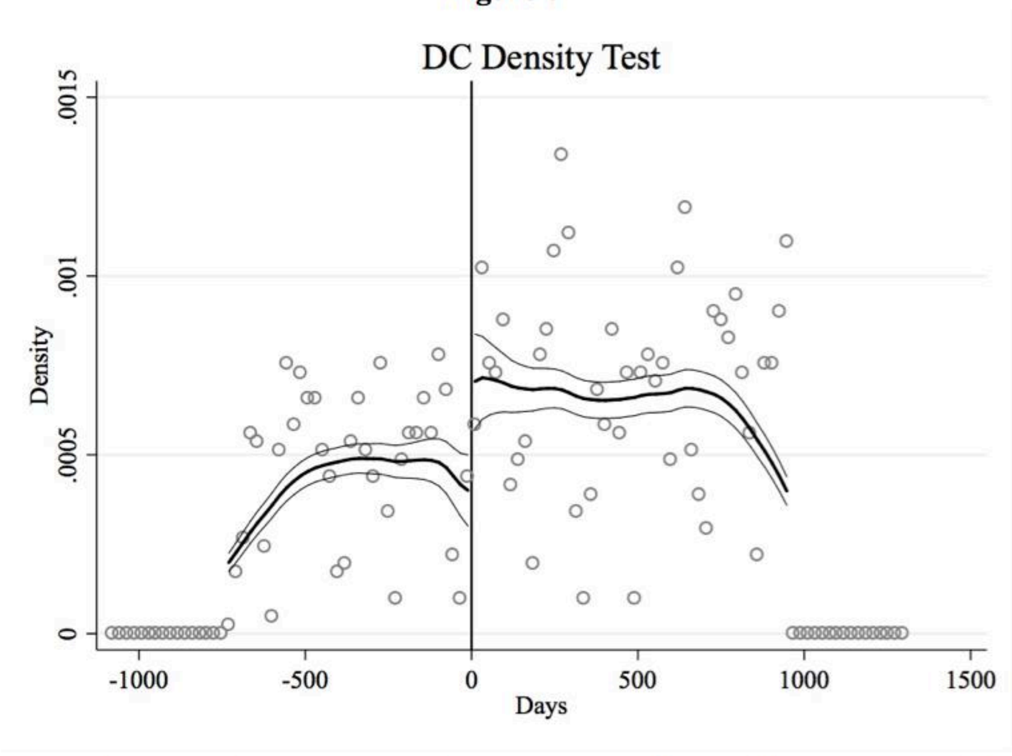
6.4 Sensitivities and Robustness Checks

6.4.1 Manipulation of running variable

If there is a possibility for individuals to manipulate the running variable (the date of the sale), the RD Design cannot correctly estimate the effect of treatment. To investigate this, we test if the density around cut-off is smooth, i.e. if there are as many observations on the left side as there are on the right side of the cut-off. Being able to manipulate the running variable would in this case be to know that there will be a price increase in properties, and based on that choose to purchase an apartment before the day of the change. Since there was no marketing leading up to the change, we believe this is highly unlikely in our case. Nevertheless, we run McCrary's (2008) density test for a discontinuity in the density of the running variable.

Figure 7 shows a visible discontinuity around cut-off: the lines are not smooth around August 18th, 2014. In spite of this, we do not believe this is due to manipulation in the assignment variable. A more reasonable explanation is the seasonal variety, illustrated in figure A3 in Appendix, since there are much fewer apartments sold during the summer, than after the end of August. This could, among other things, be a reason for the non- or low significance of our results, and the difference between global and local estimations.

Figure 7



Note: This figure shows the DC Density test of our assignment variable days, following McCrary (2007).

6.4.2 Discontinuity and Non-Linearity

In the local linear RD (Tables 6 and A8) we include several different bandwidths and polynomial orders. It is clear that the results are not robust across a wide range of polynomials. Both the standard errors and point estimates change when the polynomial of order changes. Neither are the results robust when the bandwidth varies. These are issues that make it difficult to distinguish between discontinuity and non-linearity, which in turn lead to insecure results.

6.5 Further Limitations

If the change would have involved a larger number of bus lines, the data set would have been larger and the results plausibly more reliable. Capturing the environmental effect of the public transportation systems' eco initiatives is difficult since they tend to evolve gradually over a long period of time; they do not have the resources to immediately switch all vehicles.

A common difficulty in RD analysis is to specify the underlying time trend. If improperly chosen, it can result in problems in terms of bias. If there are several variables with different time trends, the specification is even more challenging. Our benchmark BIC provides little guidance in choosing which the most appropriate trend is. A reason for this could be that other variables in the hedonic model capture much of the time trend.

Moreover, another challenging part of this study is the construction of a hedonic model with high explanatory value. Even though we find determining factors for pricing apartments in central Stockholm, the unexplained part (i.e. the residuals) might be correlated with the time trend. If that is the case, the residuals are also correlated with 1(BDH), which causes biased estimates of the coefficient for 1(BDH). As a result, this leaves the estimates from the RD Design as well.

7 Conclusion and Implications

It is generally agreed that people of today tend to be more aware of the negative effects of a bad environment. Variables such as air and noise pollution are factors that seem to be important. The purpose of this thesis is to examine to what extent this is valid. The question we have lifted is whether value is created by the introduction of more environmental-friendly buses in the neighborhood. In other words: Does closeness to a bus stop trafficked by green buses create a rise in property prices or are other factors more important?

By using a highly justified hedonic pricing model we are able to control for most factors that people value when they purchase an apartment. Thereafter, we analyze the residuals by performing a RD Design analysis.

The global RD estimates, making use of the entire data set, provide results of a price increase of between 150,000 and 190,000 SEK in properties as a result of the introduction of more environmental-friendly buses. The local method however, does not support the findings in global RD, indicating little evidence that risen prices on property are caused by the switch to cleaner fuel. Furthermore, we compare apartments located closer and farther away from the bus stops. However, our hypothesis that closer apartments are affected more is not empirically supported. Moreover, whether the air and noise improvements went unnoticed, or that they are not incorporated in housing prices is difficult to say. A larger and more evenly distributed sample would help bring clarity to this.

Yet, although our study present limited empirical evidence of an improvement in the local environment through the change to cleaner fuels, the potential benefits from improved air quality are large. Extensive research show the effect of pollutants on child mortality. Again, if we are able to incorporate both health and non-health effects, the value of environmental objectives can become more graspable for policymakers. The realization that this technology, besides being good for the environment, also brings monetary value can provide policymakers with further incentive to take measures in order to create a cleaner environment for people to live in and children to grow up in.

Finally, it remains to be seen whether so called green buses create value to properties or not. Our study has not given a clear answer to the question asked. This might be explained by the fact that our data was insufficient. On the other hand, the answer could simply be, that closeness to some kind of public transportation is important whereas the fuel used is of less importance and consequently does not affect property prices. Today we cannot present a clear theory but perhaps the future holds an answer to the question.

References

AB Storstockholms Lokaltrafik, 2016. *Fakta om SL och länet 2015,* Stockholm: AB Storstockholms Lokaltrafik.

Ashenfelter, O., 2006. MEASURING THE VALUE OF A STATISTICAL LIFE: PROBLEMS AND PROSPECTS, Cambridge: NATIONAL BUREAU OF ECONOMIC RESEARCH.

Bajic, V., 1983. The effects of a new subway line on housing prices in metropolitan Toronto. *Urban Studies*, 20(2), pp. 147-158.

Baum-Snow, N. & Kahn, M. E., 2000. The effects of new public projects to expand urban rail transit.. *Journal of Public Economics*, 77(2), p. 241.

Bento, A., Kaffine, D., Roth, K. & Zaragoza-Watkins, M., 2011. The Effects of Regulation in the Presence of Multiple Unpriced Externalities: Evidence from the Transportation Sector. *American Economic Journal: Economic Policy*, 6(3), pp. 1-29.

Booli Search Technologies AB, 2016. Så dyra är bostäderna på din tunnelbanestation - Booli. [Online]
Available at: http://www.mynewsdesk.com/se/booli.se/pressreleases/saa-dyra-aer-bostaederna-paa-din-tunnelbanestation-1393929
[Accessed 15 04 2017].

Bowes, D. R. & Ihlanfeldt, K. R., 2001. Identifying the Impacts of Rail Transit Stations on Residential Property Values. *Journal of Urban Economics*, 50(1), pp. 1-25.

Brewer, M., 2015. *Youtube*. [Online]
Available at: https://www.youtube.com/watch?v=3FTwlRFwdfI
[Accessed 1 April 2017].

Böhlin, J., 2015. *Trafikförvaltningens hållbarhetsarbete i samband med trafikupphandling*. [Online]
Available at: http://www.lansstyrelsen.se/stockholm/SiteCollectionDocuments/Sv/miljo-och-klimat/klimat-och-energi/Klimat%20och%20energi/Seminarier/Presentationer%20seminarium%2026%20maj%202015/Presentation%20Johan%20B%C3%B6hlin,%20SLL%20Trafikf%C3%B6rvaltningen.pdf
[Accessed 10 April 2017].

Calonico, S., Cattaneo, M. & Titiunik, R., 2014. Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*, 14(4), pp. 909-946.

Calonico, S., Cattaneo, M., Titiunik, R. & Farrell, M., 2017. rdrobust: Software for Regression Discontinuity Designs. *The Stata Journal*, 18 January, pp. 1-33.

Can, A., 1992. Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, 22, pp. 453-474.

Cervero, R. & Kang, C. D., 2011. Bus rapid transit impacts on land uses and land values in Seoul, Korea. *Transport Policy*, 18(1), pp. 102-116.

Chen, Y. & Whalley, A., 2012. Green Infrastructure: The Effects of Urban Rail Transit on Air Quality. *American Economic Journal: Economic Policy*, 4(1), pp. 58-97.

Chin, T. & Chau, K., 2003. A Critical Review of Literature on the Hedonic Price Model. *International Journal for Housing Science and Its Applications,* 2(27), pp. 145-165.

Clay, K., Lewis, J. & Severnini, E., 2016. *Canary in a Coal Mine: Infant Mortality, Property Values, and Tradeoffs Associated with Mid-20th Century Air Pollution,* Cambridge: NATIONAL BUREAU OF ECONOMIC RESEARCH.

Concas, S., 2013. Accessibility and housing price resilience: Evidence from limited-access roadways in Florida. *Transportation Research Record: Journal of the Transportation Research Board*, 2357, pp. 66-76.

Damm, D., Lerner-Lam, E. & Young, J., 1980. Response of Urban Real Estate Values in Anticipation of the Washington Metro. *Journal of Transport Economics and Policy*, 14(3), pp. 315-336.

Davis, L., 2008. The Effect of Driving Restrictions on Air Quality in Mexico City. *Journal of Political Economy*, 116(1), pp. 38-81.

Des Rosiers, F., Thériault, M., Voisin, M. & Dubé, J., 2010. Does an Improved Urban Bus Service Affect House Values?. *International Journal of Sustainable Transportation*, 4(6), pp. 321-346.

Dewees, D. N., 1976. The Effect of a Subway on Residential Property Values in Toronto. *Journal of Urban Economics*, 3(4), p. 357.

Dube, J., Rosiers, F. D., Theriault, M. & Dib, P., 2011. Economic impact of a supply change in mass transit in urban areas: A Canadian example. *Transportation Research Part A*, 45(1), pp. 46-62.

Dubin, R., 1998. Predicting house prices using multiple listings data. *The Journal of Real Estate Finance and Economics*, 17(1), pp. 35-60.

Ecotraffic, 2015. *Stadsbussar - Kunskapssammanställning EURO VI,* u.o.: Ecotraffic.
Fama, E., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance,* pp. 383-417.

Greenstone, M. & Chay, K., 2003. The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *The Quarterly Journal of Economics,* 118(3), pp. 1121-1167.

Hickman, E., Gaines, J. & FJ, I., 1984. The influence of neighborhood quality on residential values. *The Real Estate Appraiser and Analyst*, 50(2), pp. 36-42.

Icopal, 2017. *Nitrogen Oxide (NOx) Pollution*. [Online]
Available at: http://www.icopal-noxite.co.uk/nox-problem/nox-pollution.aspx
[Accessed 9 Maj 2017].

Krantz, D., Waever, R. & TR, A., 1982. Residential tax capitalization: Consistent estimates using micro-level data. *Land Economics*, 58(4), pp. 488-496.

Länsstyrelsen Blekinge Län, 2005. *VOC- lättflyktiga organiska ämnen (Volatile Organic Compounds)*.. [Online]
Available at: http://www.lansstyrelsen.se/blekinge/Sv/miljo-och-klimat/verksamheter-med-miljopaverkan/miljofarlig-verksamhet/Pages/voc-direktivet.aspx
[Accessed 9 Maj 2017].

Länsstyrelsen Stockholm, u.d. *Bullernätverket Stockholms Län*. [Online]
Available at: http://www.bullernatverket.se/kartlaggning-2/
[Accessed 10 05 2017].

Lee, D. & Lemieux, T., 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, June, 48, pp. 281-355.

Maasing, U., 2013. *Nästan 200 MAN-bussar till Keolis i Stockholm*. [Online]
Available at: http://www.bussmagasinet.se/2013/12/nastan-200-man-bussar-till-keolis-i-stockholm/
[Accessed 20 Mars 2017].

McCrary, J., 2008. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142, pp. 698-714.

McDonald, J. F. & Osuji, C. I., 1995. The effect of anticipated transportation improvement on residential land values. *Regional Science & Urban Economics*, 25(3), p. 261.

McLeod, P., 1984. The demand for local amenity: an hedonic price analysis. *Environment & Planning A*, 16(3), pp. 389-400.

McMillen, D. & McDonald, J., 2004. Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983–1999. *Real Estate Economics*, 32(3), pp. 463-486.

Mikelbank, B. A., 2004. Spatial analysis of the relationship between housing values and investments in transportation infrastructure. *Annals of Regional Science*, 38(4), pp. 705-726.

Monson, M., 2009. Valuation using hedonic pricing models. *Cornell Real Estate Review,* 7, pp. 62-73.

Mulley, C. & Tsai, C., 2017. Impact of bus rapid transit on housing price and accessibility changes in Sydney: A repeat sales approach. *International Journal of Sustainable Transportation*, 11(1), pp. 3-10.

Naturvårdsverket, 2017. *Nitrogen Oxide (NOx) Pollution - Health Issues - Icopal*. [Online]
Available at: http://www.naturvardsverket.se/Sa-mar-miljon/Manniska/Buller/
[Accessed 10 May 2017].

Orton, K., 2017. Stockholm. Reference at Keolis AB.

Raftery, A., 1995. Bayesian Model Selection in Social Research. *Sociological Methodology*, 25, pp. 111-163.

Rodriguez, D. A. & Mojica, C. H., 2009. Capitalization of BRT network expansions effects into prices of non-expansion areas. *Transportation Research Part A*, 43(5), pp. 560-571.

Shefer, D., 1986. Utility changes in housing and neighborhood services for households moving into and out of distressed neighborhoods. *Journal of Urban Economics*, 19, pp. 107-124.

Stockholms stad, 2014. Luften i Stockholm Årsrapport, Stockholm.

Strange, W., 1992. Overlapping neighborhoods and housing externalities. *Journal of Urban Economics*, 32, pp. 17-39.

Thériault, M., Des Rosiers, F. & Dubé, . J., 2007. Testing the temporal stability of accessibility value in residential hedonic prices. *Scienze Regionali*, 6(3), pp. 5-46.

Thistlethwaite, D. & Campbell, D., 1960. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6), pp. 309-317.

Trafikförvaltningen Stockholms läns landsting, 2014. *Delårsrapport augusti 2014 för Trafiknämnden,* Stockholm: Trafikförvaltningen Stockholms läns landsting.

Vargiu, E. & Urru, M., 2013. Exploiting web scraping in a collaborative filtering- based approach to web advertising. *Artificial Intelligence Research,* 2(1).

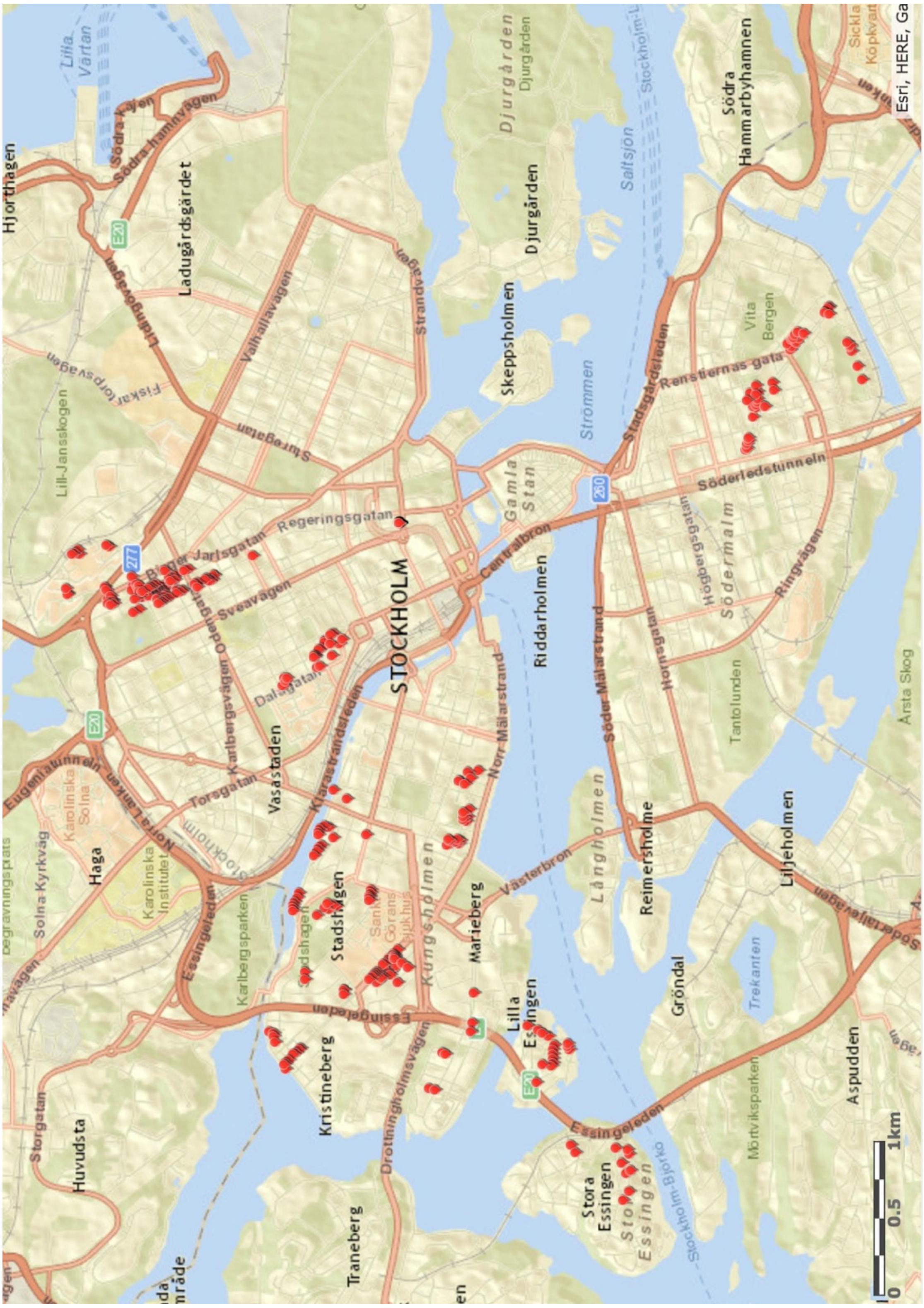
Wilhelmsson, M., 2000. The impact of traffic noise on the values of single family houses. *Journal of Environmental Planning and Management*, 43(6), pp. 799-815.

Viscusi, K. & Aldy, J., 2003. *THE VALUE OF A STATISTICAL LIFE: A CRITICAL REVIEW OF MARKET ESTIMATES THROUGHOUT THE WORLD,* Cambridge: NATIONAL BUREAU OF ECONOMIC RESEARCH.

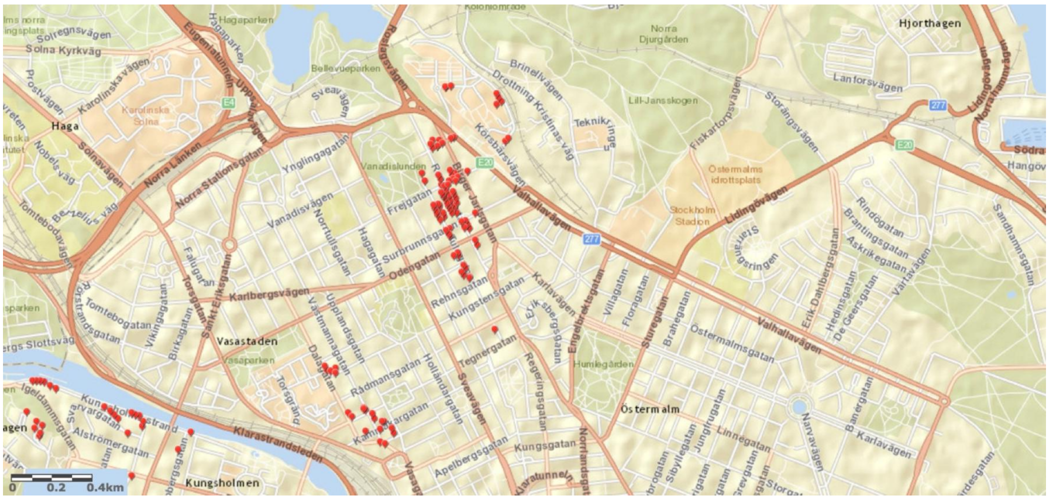
Voith, R., 1993. Changing Capitalization of CBD-Oriented Transportation Systems: Evidence from Philadelphia, 1970-1988. *Journal of Urban Economics*, 33(3), p. 361.

Appendix

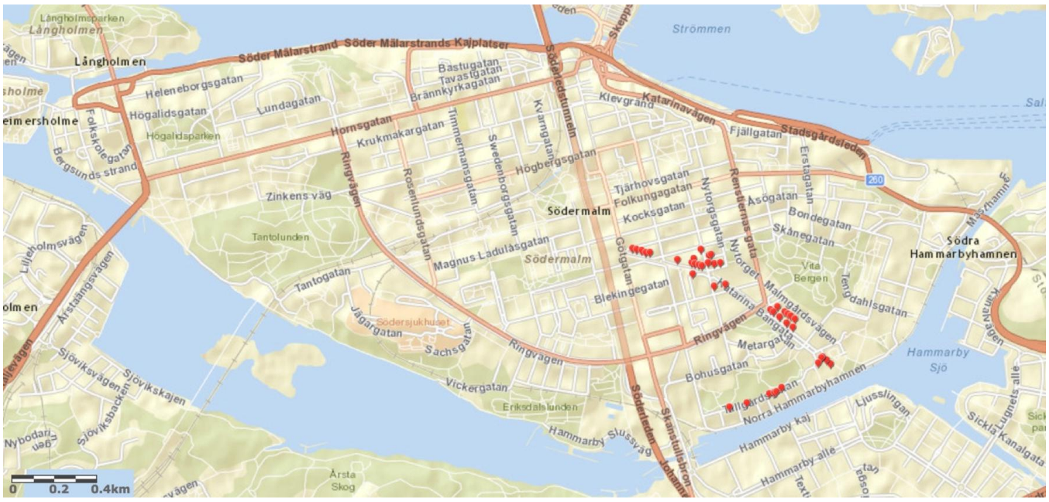
A1



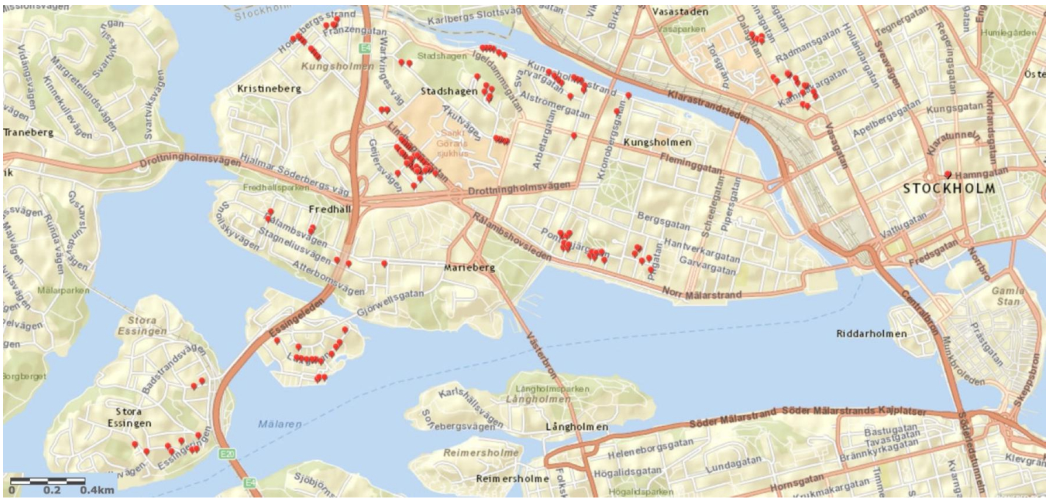
Note: All observations in Stockholm



Note: Observations located in Vasastan



Note: Observations located on Södermalm



Note: Observations located on Kungsholmen, Lilla Essingen and Stora Essingen

A2

Table A2

Equation	(1)	(2)	(3)	(4)
VARIABLES	price	price	price	price
1.(BDH)	253,339**	230,676**	175,910**	175,559**
	(118,238)	(116,239)	(86,666)	(86,546)
space	66,532***	66,390***	35,055***	34,845***
	(2,843)	(2,804)	(6,001)	(5,654)
fee	-198.9***	-198.4***	-169.7*	-163.3***
	(42.33)	(42.19)	(95.10)	(38.78)
fee2			0.00101	
			(0.0157)	
wdm_x	155,354**	155,200**	172,064***	171,997***
	(72,424)	(72,763)	(62,761)	(62,841)
space2			249.6***	251.1***
			(50.93)	(47.29)
rooms	187,836***	189,575***		
	(44,996)	(44,407)		
central	453,693***	473,387***		
	(110,808)	(110,153)		
floor	67,239***			
	(12,534)			
2.floor_int		129,127***	116,520***	116,348***
		(38,496)	(36,965)	(36,719)
3.floor_int		150,742***	140,938***	140,713***
		(46,378)	(43,733)	(44,010)
4.floor_int		193,100***	193,902***	193,663***
		(48,509)	(45,593)	(45,438)
5.floor_int		241,587***	236,916***	236,582***
		(70,822)	(69,355)	(69,091)
6.floor_int		508,957***	460,443***	460,166***
		(82,236)	(82,804)	(83,105)
7.floor_int		404,728***	404,527***	404,515***
		(109,233)	(107,225)	(107,120)
8.floor_int		474,176***	361,913**	362,501**
		(176,839)	(169,967)	(169,275)
9.floor_int		320,883*	276,470	276,431
		(170,472)	(186,134)	(186,195)
10.floor_int		577,880	512,121	511,830
		(402,016)	(444,199)	(444,230)
1.central			365,424***	366,142***
			(105,664)	(104,767)
2.rooms_int			158,542***	158,619***
			(36,634)	(36,551)
3.rooms_int			395,618***	395,978***
			(81,014)	(81,279)
4.rooms_int			432,012***	433,416***
			(165,001)	(160,676)
5.rooms_int			-98,071	-94,186
			(366,864)	(365,433)
2.qq			73,233*	73,164*
			(43,522)	(43,359)
3.qq			944,366***	944,306***
			(143,711)	(143,704)
4.qq			950,171***	950,078***
			(149,763)	(149,650)
Constant	78,035	140,345	732,303***	729,792***
	(135,170)	(132,637)	(160,789)	(152,834)
Observations	1,617	1,617	1,617	1,617
R-squared	0.945	0.945	0.949	0.949
Adj. R-Squared	0.930	0.930	0.935	0.935

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table reports the estimates from four separate regressions of equations (1), (2), (3) and (4), that are made in order to capture the best fitting hedonic pricing model for the data set. The dependent variable is the property price in SEK (collected April 1st, 2017). The coefficients of the regression variables should be interpreted in monetary terms, SEK. Standard errors are in parentheses and P-values of the OLS are in square brackets. R-squared ranges from 0.945 to 0.949, and adjusted R-squared ranges from 0.930 to 0.935.

A3

Table A3(a)

Year	Freq.
2012	76
2013	347
2014	391
2015	440
2016	490
2017	144
Total	**1 888**

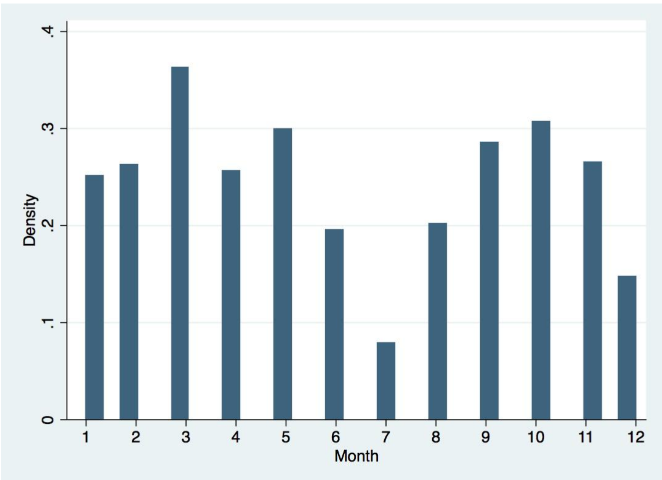
Note: This table shows the distribution of transactions across years

Table A3(b)

Month	Freq.	Percent
1	163	8.63
2	170	9.00
3	235	12.45
4	166	8.79
5	194	10.28
6	127	6.73
7	51	2.70
8	131	6.94
9	185	9.80
10	199	10.54
11	172	9.11
12	95	5.03
Total	**1,888**	**100.00**

Note: This table shows the distribution of transactions across months

Figure A3



Note: Monthly transactions of properties during 2012 to 2017.

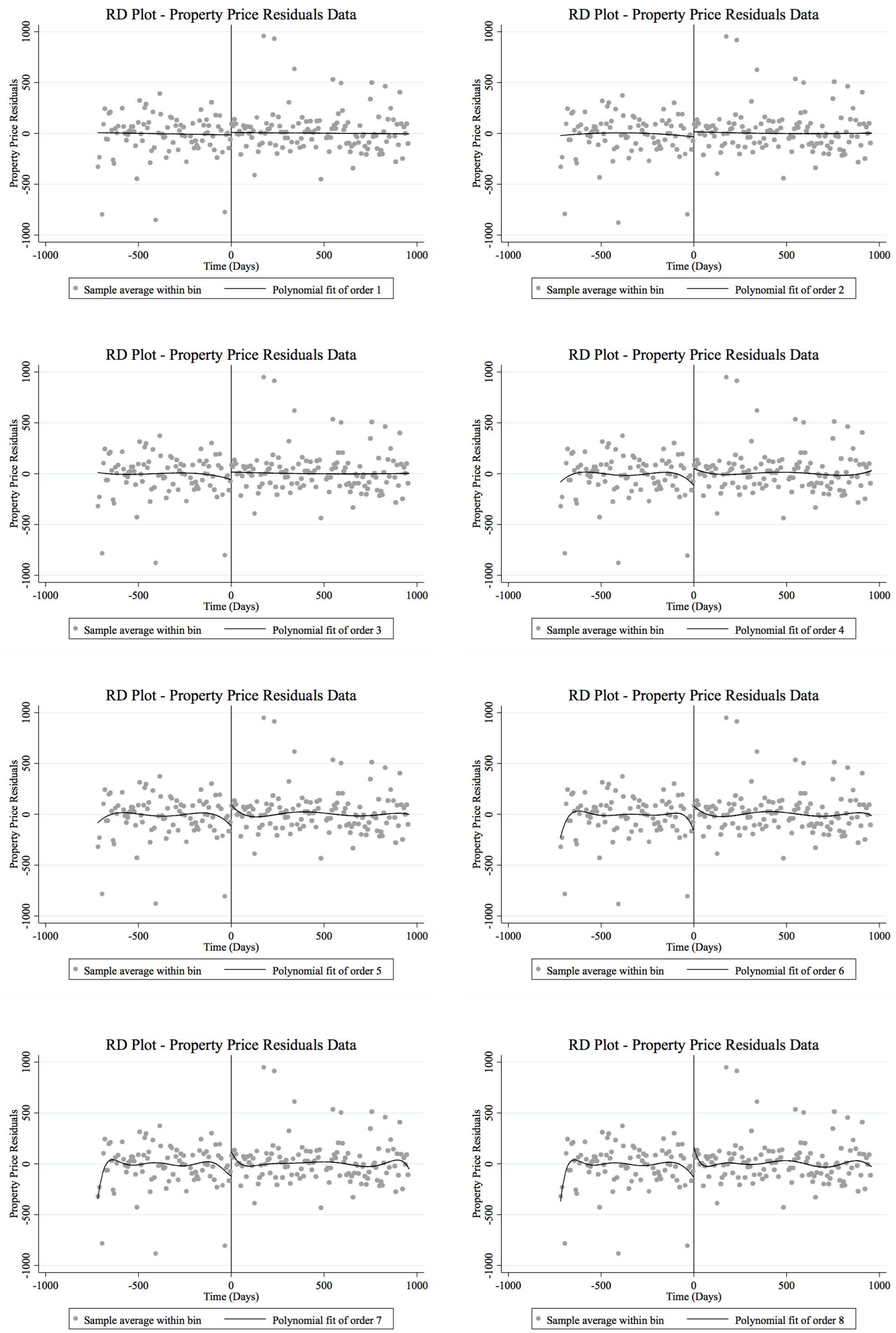
A4

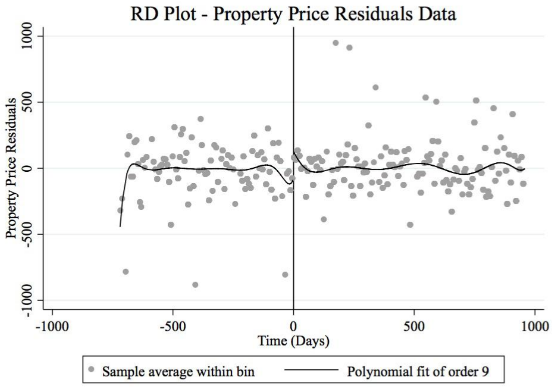
Table A4

Zip Code	Half-Year										
	2012,2	2013,1	2013,2	2014,1	2014,2	2015,1	2015,2	2016,1	2016,2	2017,1	**Total**
11124	1	5	1	5	7	2	2	3	2	1	29
11215		1				2	1			1	5
11216	3	3	2	1	3	8	4	3	3	2	32
11218			1			4		1	4	2	12
11222	2	8	4	3	5	3	5	8	6	3	47
11223		2									2
11233	1			3	3	1		3		1	12
11235		1			1		1	4		2	9
11236	1	3	3	2	1	3	3	1	1	1	19
11237	3	6	6	5	3	2	6	5	10	5	51
11239								2		1	3
11240							1				1
11243	9	21	17	21	20	33	13	28	22	17	201
11244	5	5	6	3	8	5	7	5	4	2	50
11245			1	2		1		3	4		11
11246	4	9	9	11	5	12	13	8	13	5	89
11248	2	9	8	10	9	13	8	13	10	5	87
11250	1	8	5	7	12	6	7	11	5	5	67
11251	4	6	7	16	20	15	12	20	20	13	133
11254	1	4	2	2	1	2	1	1	3	1	18
11255		3	1	3	5	6	1	5	1	2	27
11256		2	2	1	3	5	3	4	9	3	32
11258									3	2	5
11259	1	2	2	1	3	5	3	6	3	3	29
11262	7	7	7	3	11	14	8	8	9	5	79
11263	7	16	7	21	11	24	9	23	16	9	143
11264		2	1	1	1		1	2		1	9
11266		1		1	3	2		3	1	2	13
11267	1	1		1			1				4
11269	1										1
11324		1	1		1				1		4
11325			1		2	1	2	3			9
11326								1	1		2
11329			2	1	2			2	2	1	10
11348		1									1
11349	1	2	3	2	5	8	3		1	2	27
11353	5	12	4	11	6	19	12	17	14	9	109
11354	1	2	5	6	5	6	3	3	8	3	42
11355	1	8	12	11	17	24	16	19	16	10	134
11358				1	1		1			2	5
11420	1	1	1	2	1	1	1		2	3	13
11421	3	8	2	4	3	8	3	3	3	2	39
11423		4	2		3	3	1	2	1	3	19
11639	1	7	4	6	6	7	2	4	4	5	46
11640	2	8	6	5	4	6	2	4	6	2	45
11642	6	12	10	13	10	13	13	24	19	8	128
11643		4	1	1	1	1		2	2		12
11668				1			1		1		3
16668	1	4	2	1	1	4	1	1	5		20
Total	**76**	**199**	**148**	**188**	**203**	**269**	**171**	**255**	**235**	**144**	**1888**

Note: This table reports the number of transactions per zip code each half year during the years 2012 to 2017, showing a relatively smooth distribution.

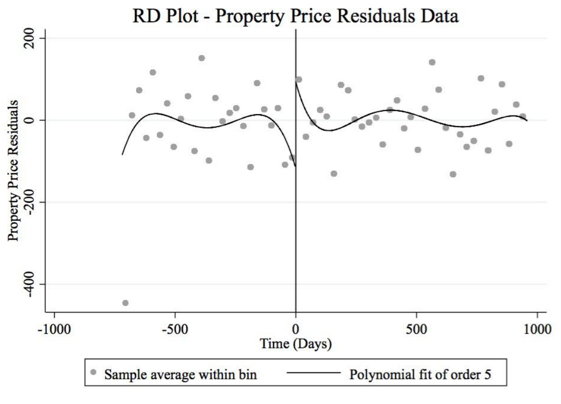
A5


RD Plot - Property Price Residuals Data
Property Price Residuals
-1000
-500
0
500
1000
Time (Days)
Sample average within bin
Polynomial fit of order 1
RD Plot - Property Price Residuals Data
Sample average within bin
Polynomial fit of order 2
RD Plot - Property Price Residuals Data
Sample average within bin
Polynomial fit of order 3
RD Plot - Property Price Residuals Data
Sample average within bin
Polynomial fit of order 4
RD Plot - Property Price Residuals Data
Sample average within bin
Polynomial fit of order 5
RD Plot - Property Price Residuals Data
Sample average within bin
Polynomial fit of order 6
RD Plot - Property Price Residuals Data
Sample average within bin
Polynomial fit of order 7
RD Plot - Property Price Residuals Data
Sample average within bin
Polynomial fit of order 8

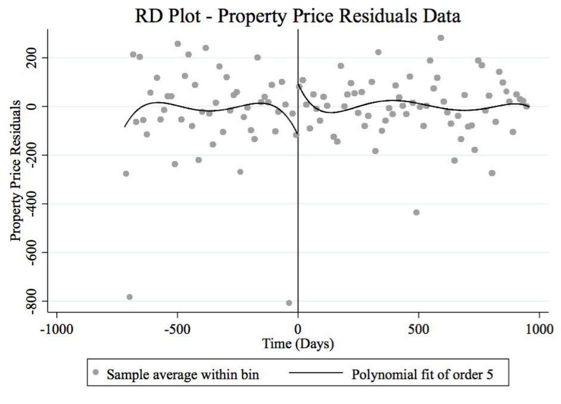


Note: This figure plots the residuals (in thousands SEK) from regression of equation (4), excluding 1(BDH), including each of the first, second, third, fourth, fifth, sixth, seventh, eighth and ninth order of polynomial time trend, respectively. Each bin is 7 days.

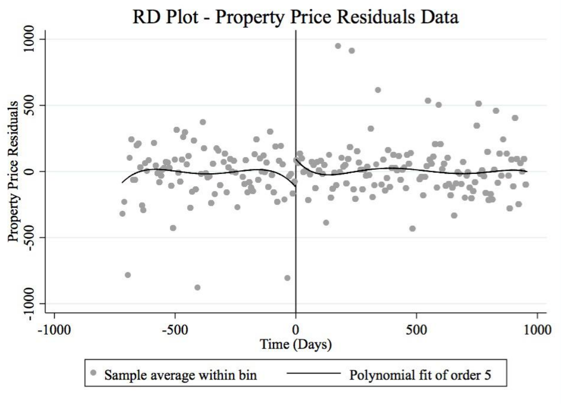
A6



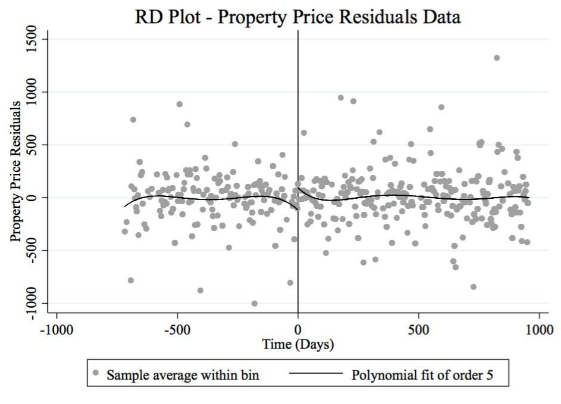
Bin length: 29 days



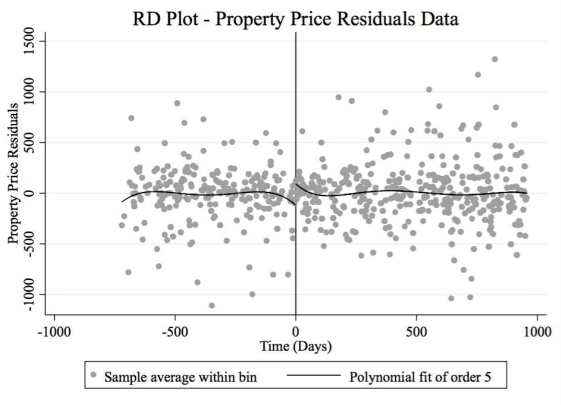
Bin length: 14 days



Bin length: 7 days



Bin length: 3.6 days



Bin length: 1.8 days

Note: This figure plots the residuals (in thousands SEK) from regression of equation (4), excluding 1(BDH), including a fifth order of polynomial time trend. The base scenario, where the bin length is 7 days, is halved and doubled in order to illustrate the

A7

Table A7 (a)

Model	(1)	(2)	(3)	(4)	(5)
VARIABLES	price	price	price	price	price
1.(BDH)	162,740	175,329	179,657	181,167	182,265
	(184,408)	(189,496)	(190,281)	(190,023)	(189,685)
Constant	1,086,000	1,315,000***	1,388,000***	1,407,000***	1,412,000***
	(865,396)	(474,312)	(358,443)	(316,100)	(296,770)
Polynomial Time Trend	1	2	3	4	5
Observations	760	760	760	760	760
R-squared	0.953	0.953	0.953	0.953	0.953
Adj. R-squared	0.935	0.935	0.935	0.935	0.935
Model	(6)	(7)	(8)	(9)	
VARIABLES	price	price	price	price	
1.(BDH)	183,138	183,698	183,907	183,787	
	(189,452)	(189,330)	(189,288)	(189,297)	
Constant	1,414,000***	1,415,000***	1,416,000***	1,418,000***	
	(286,642)	(280,792)	(277,111)	(274,587)	
Polynomial Time Trend	6	7	8	9	
Observations	760	760	760	760	
R-squared	0.953	0.953	0.953	0.953	
Adj. R-squared	0.935	0.935	0.935	0.935	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table reports global estimates from nine separate regressions of equation (4) for properties located directly by bus stop, excluding 1(BDH) and including each of the first, second, third, fourth, fifth, sixth, seventh, eighth and ninth order of the time trend, respectively. R-squared is 0.953 and adjusted R-squared is 0.935 for all nine regressions.

Table A7 (b)

Model	(1)	(2)	(3)	(4)	(5)
VARIABLES	price	price	price	price	price
1.(BDH)	177,380*	186,171**	185,589**	185,957**	186,859**
	(94,986)	(92,109)	(91,829)	(92,179)	(92,427)
Constant	541,804	769,837**	767,800***	759,323***	753,475***
	(513,621)	(314,638)	(269,651)	(257,064)	(252,583)
Polynomial Time Trend	1	2	3	4	5
Observations	857	857	857	857	857
R-squared	0.959	0.959	0.959	0.959	0.959
Adj. R-squared	0.943	0.943	0.943	0.943	0.943
Model	(6)	(7)	(8)	(9)	
VARIABLES	price	price	price	price	
1.(BDH)	187,855**	188,796**	189,637**	190,360**	
	(92,533)	(92,557)	(92,543)	(92,510)	
Constant	749,353***	746,187***	743,613***	741,472***	
	(250,844)	(250,220)	(250,093)	(250,188)	
Polynomial Time Trend	6	7	8	9	
Observations	857	857	857	857	
R-squared	0.959	0.959	0.959	0.959	
Adj. R-squared	0.943	0.943	0.943	0.943	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table) reports global estimates from nine separate regressions of equation (4) for properties located in-directly by bus stops, excluding 1(BDH) and including each of the first, second, third, fourth, fifth, sixth, seventh, eighth and ninth order of the time trend, respectively. R-squared is 0.959 and adjusted R-squared is 0.943 for all nine regressions. The coefficients of the regression variables in Table A7 (a) and (b) should be interpreted in monetary terms, SEK. Standard errors are in parentheses.

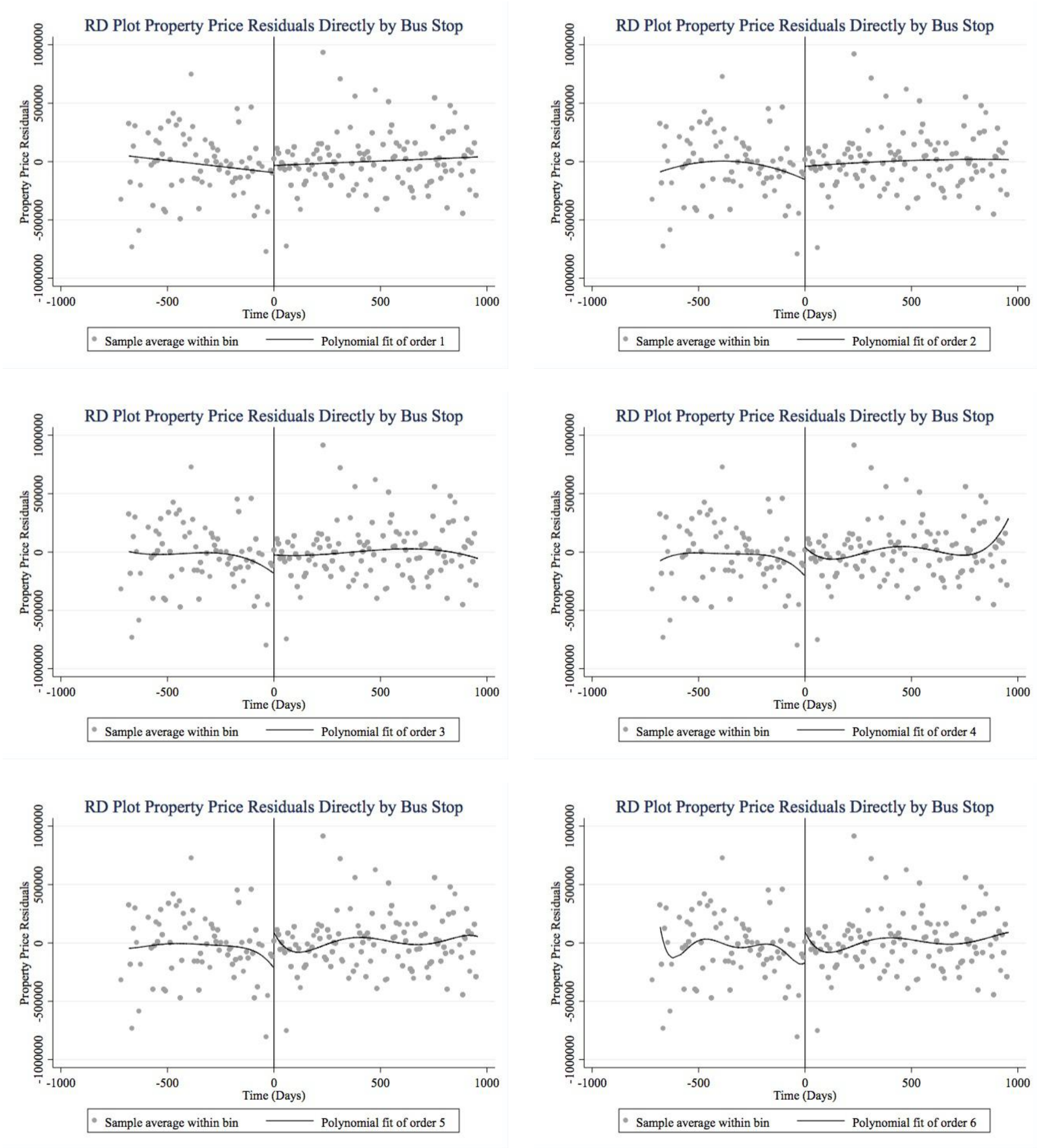
A8

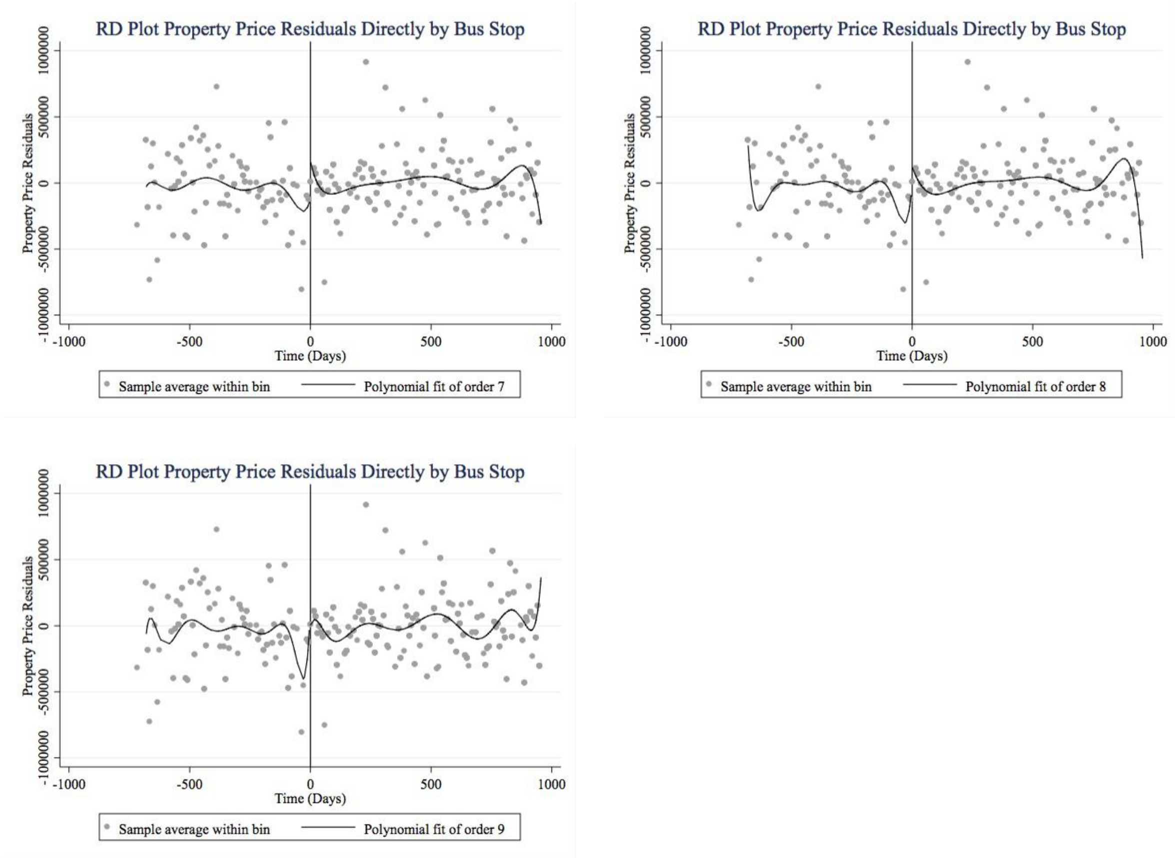
Table A8

Bandwidth	**15**		**30**		**40**	
Location:	**Direct**	**In-direct**	**Direct**	**In-direct**	**Direct**	**In-direct**
Polynomial of order:						
One	N/A	183,598	180.6	236,747*	-46,268	255,592*
		(224,172)	(245,755)	(141,003)	(213,353)	(139,154)
		[0.413]	[0.999]	[0.093]	[0.828]	[0.066]
Two	N/A	1,227,000	128,974	98,626	132,094	44,997
		(1,085,000)	(526,229)	(245,339)	(445,146)	(237,423)
		[0.258]	[0.806]	[0.688]	[0.767]	[0.850]
Three	N/A	N/A	N/A	154,100	436,482	28,517
				(513,001)	(886,175)	(490,214)
				[0.764]	[0.622]	[0.954]
Observations	*13*	*14*	*28*	*30*	*38*	*37*
Bandwidth	**50**		**60**		**70**	
Location:	**Direct**	**In-direct**	**Direct**	**In-direct**	**Direct**	**In-direct**
Polynomial of order:						
One	-31,306	254,339*	69,892	255,836**	195,969	254,530***
	(202,829)	(130,786)	(194,734)	(105,487)	(161,999)	(94,239)
	[0.877]	[0.052]	[0.720]	[0.015]	[0.226]	[0.007]
Two	134,637	233,013	-164,305	232,613	-138,207	245,302*
	(434,924)	(168,578)	(279,370)	(154,956)	(244,379)	(143,097)
	[0.757]	[0.167]	[0.556]	[0.133]	[0.572]	[0.087]
Three	386,739	-12,702	244,996	150,051	-63,392	217,486
	(850,546)	(295,025)	(558,342)	(248,975)	(383,921)	(174,042)
	[0.649]	[0.966]	[0.661]	[0.547]	[0.869]	[0.211]
Observations	*45*	*43*	*52*	*52*	*63*	*64*

Note: This table reports the local estimates from 24 separate regressions of the residuals from equation (4) over the whole data set, excluding 1(BDH), including a fourth order of polynomial time trend. Bandwidths are in number of days. The coefficients of the regression variables should be interpreted in monetary terms, SEK. Conventional standard error in parenthesis, and conventional p-value in brackets. N/A if there are too few observations to compute the estimate.

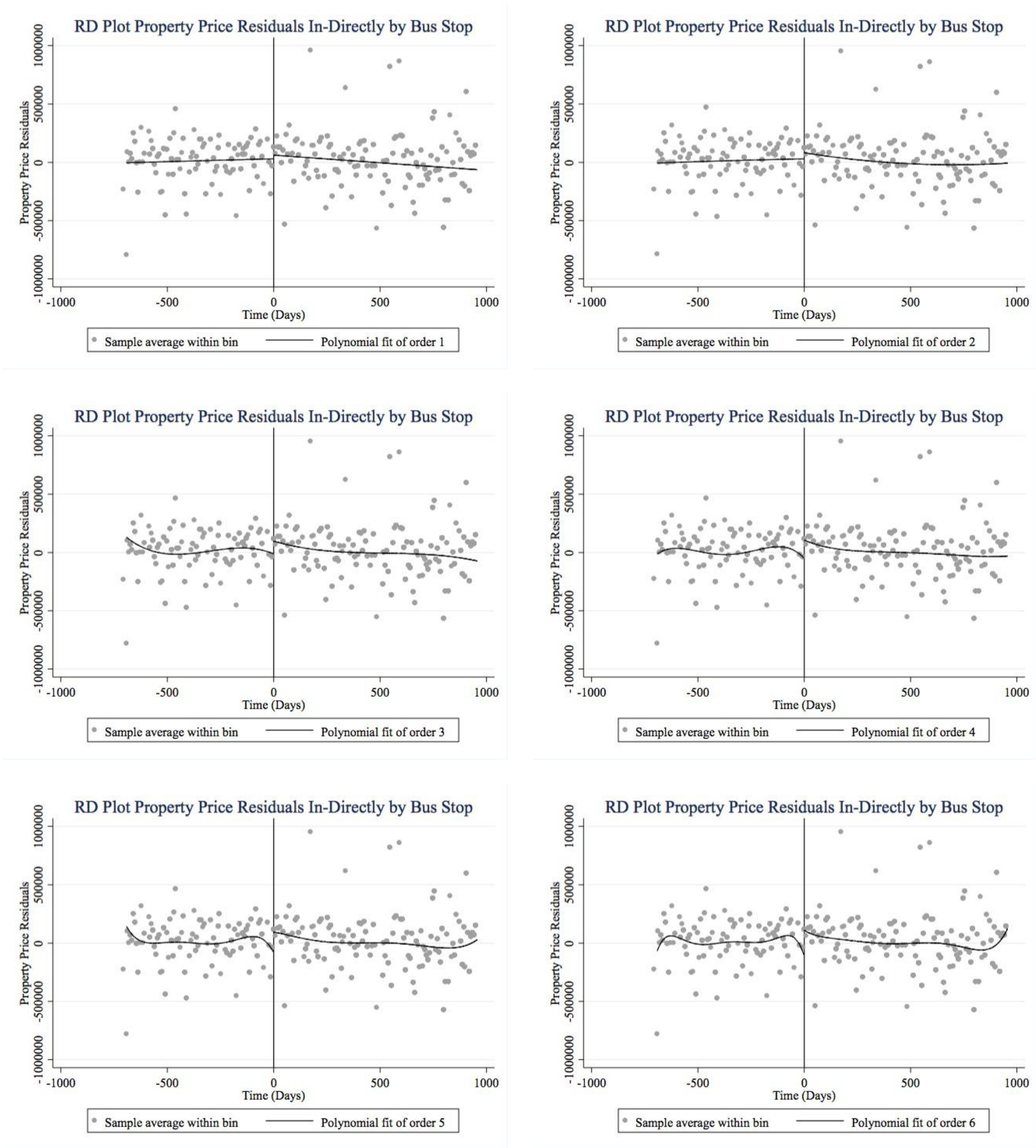
A9

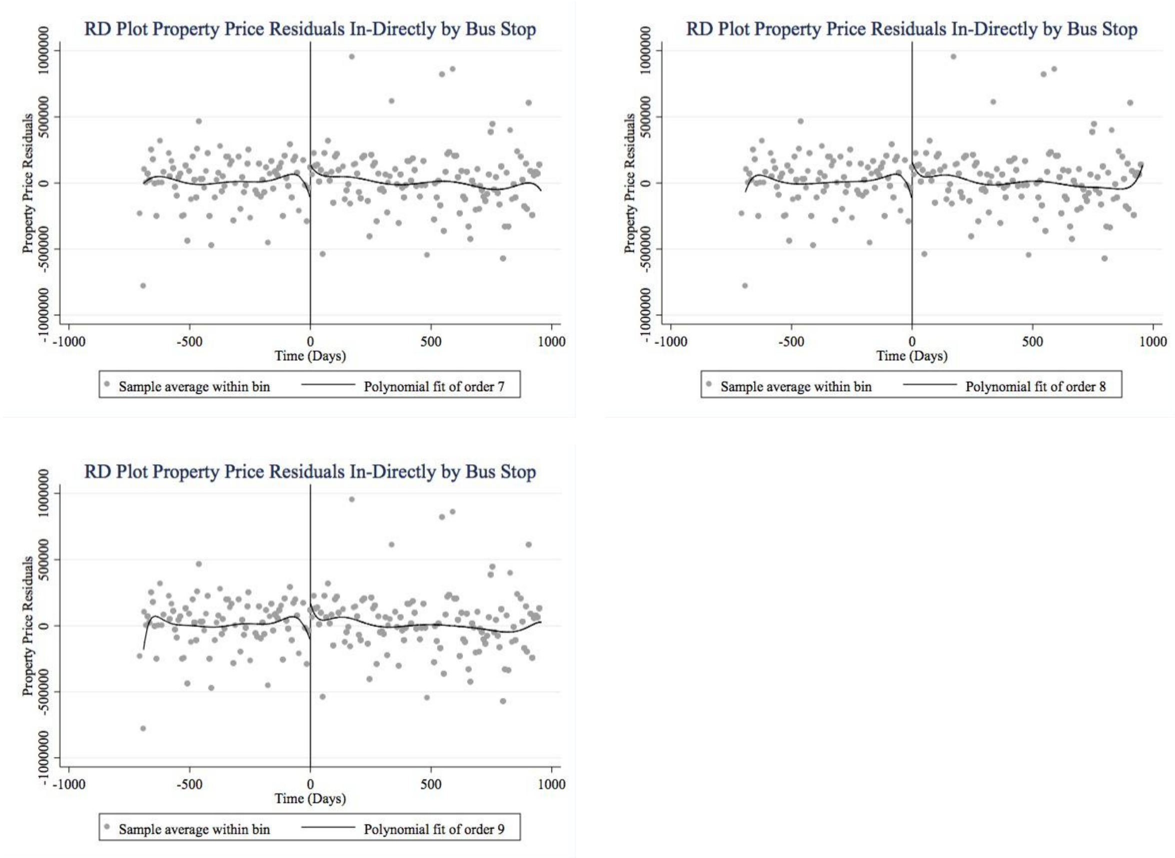

RD Plot Property Price Residuals Directly by Bus Stop
Property Price Residuals
-1000000
-500000
0
500000
1000000
-1000
-500
0
500
1000
Time (Days)
Sample average within bin
Polynomial fit of order 1
Polynomial fit of order 2
Polynomial fit of order 3
Polynomial fit of order 4
Polynomial fit of order 5
Polynomial fit of order 6



Note: This figure plots the residuals (in thousands SEK) from regression of equation (4) for properties located directly by bus stops, excluding 1(BDH), including each of the first, second, third, fourth, fifth, sixth, seventh, eighth and ninth order of polynomial time trend, respectively. Each bin is 7 days.

A10


RD Plot Property Price Residuals In-Directly by Bus Stop
Property Price Residuals
-1000000
-500000
0
500000
1000000
-1000
-500
0
500
1000
Time (Days)
Sample average within bin
Polynomial fit of order 1
RD Plot Property Price Residuals In-Directly by Bus Stop
Polynomial fit of order 2
RD Plot Property Price Residuals In-Directly by Bus Stop
Polynomial fit of order 3
RD Plot Property Price Residuals In-Directly by Bus Stop
Polynomial fit of order 4
RD Plot Property Price Residuals In-Directly by Bus Stop
Polynomial fit of order 5
RD Plot Property Price Residuals In-Directly by Bus Stop
Polynomial fit of order 6



Note: This figure plots the residuals (in thousands SEK) from regression of equation (4) for properties located in-directly by bus stops, excluding 1(BDH), including each of the first, second, third, fourth, fifth, sixth, seventh, eighth and ninth order of polynomial time trend, respectively. Each bin is 7 days.

A11

Code for extracting the data from Booli

```
using System;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
using HtmlAgilityPack;
namespace WindowsFormsApp7{
    public class HomeReview{
        public string Name { get; set; }
        public string Price { get; set; }
        public string Date { get; set; }
        public string Floor { get; set; }
        public string Fee { get; set; }
    }
    public partial class Form1 : Form
    {DataTable table;
        HtmlWeb web = new HtmlWeb();
        public Form1()
        {InitializeComponent();
            InitTable();
        }private void InitTable() {
            table = new DataTable("RealEstateDataTable");
            table.Columns.Add("Name", typeof(string));
            table.Columns.Add("Price", typeof(string));
            table.Columns.Add("Date", typeof(string));
            table.Columns.Add("Floor", typeof(string));
            table.Columns.Add("Fee", typeof(string));
            RealEstateDataView.DataSource = table;}
        private async Task<List<HomeReview>> RealEstateFromPage(int pageNum) {
            string url = "https://www.booli.se/slutpriser/stockholms+innerstad/143/";
            if (pageNum != 0)
                url = "https://www.booli.se/slutpriser/stockholms+innerstad/143/?page=" +
pageNum.ToString();
            var doc = await Task.Factory.StartNew(() => web.Load(url));
            var nameNodes = doc.DocumentNode.SelectNodes("//*[@id=\"js__search-
list\"]/ul//li//a/span[2]/span[1]");
            var priceNodes = doc.DocumentNode.SelectNodes("//*[@id=\"js__search-
list\"]/ul//li//a/span[3]/span[1]");
            var dateNodes = doc.DocumentNode.SelectNodes("//*[@id=\"js__search-
list\"]/ul//li//a/span[3]/span[3]");
            var floorNodes = doc.DocumentNode.SelectNodes("//*[@id=\"js__search-
list\"]/ul//li//a/span[2]/span[2]");
            var feeNodes = doc.DocumentNode.SelectNodes("//*[@id=\"js__search-
list\"]/ul//li//a/span[4]/span[2]");
            //If these are null it means the nodes couldn't be found on the html page
            if (nameNodes == null || priceNodes == null || dateNodes == null || floorNodes == null ||
feeNodes == null)
                return new List<HomeReview>();
            var names = nameNodes.Select(node => node.InnerText).ToList();
            var prices = priceNodes.Select(node => node.InnerText).ToList();
            var dates = dateNodes.Select(node => node.LastChild.InnerText).ToList();
            var floors = floorNodes.Select(node => node.InnerText).ToList();
            var fee = feeNodes.Select(node => node.InnerText).ToList();
            List<HomeReview> toReturn = new List<HomeReview>();
            for (int i = 0; i < names.Count(); ++i)
                toReturn.Add(new HomeReview() { Name = names[i], Price = prices[i], Date = dates[i],
Floor = floors[i] , Fee = fee[i] });
           return toReturn; }
        private async void Form1_Load(object sender, EventArgs e) {
            int pageNum = 0;
            var rankings = await RealEstateFromPage(0);
            while (rankings.Count > 0) {
                foreach (var ranking in rankings)
                    table.Rows.Add(ranking.Name, ranking.Price, ranking.Date, ranking.Floor,
ranking.Fee);
                pageNum++;
                rankings = await RealEstateFromPage(pageNum); }
```