Do segregated housing markets have a spillover effect on housing prices in nearby residential areas?

Mohammad Ismail, Abukar Warsame and Mats Wilhelmsson
Department of Real Estate Management, Royal Institute of Technology, Stockholm, Sweden

Abstract

Purpose – The purpose of this study is to analyse the trends regarding housing segregation over the past 10–20 years and determine whether housing segregation has a spillover effect on neighbouring housing areas. Namely, the authors set out to determine whether proximity to a specific type of segregated housing market has a negative impact on nearby housing markets while proximity to another type of segregated market has a positive impact.

Design/methodology/approach – For the purposes of this paper, the authors must combine information on segregation within a city with information on property values in the city. The authors have, therefore, used data on the income of the population and data on housing values taken from housing transactions. The case study used is the city of Stockholm, the capital of Sweden. The empirical analysis will be the estimation of the traditional hedonic pricing model. It will be estimated for the condominium market.

Findings – The results indicate that segregation, when measured as income sorting, has increased over time in some of the housing markets. Its effects on housing values in neighbouring housing areas are significant and statistically significant.

Research limitations/implications – A better understanding of the different potential spillover effects on housing prices in relation to the spatial distribution of various income groups would be beneficial in determining appropriate property assessment levels. In other words, awareness of this spillover effect could improve existing property assessment methods and provide local governments with extra information to make an informed decision on policies and services needed in different neighbourhoods.

Practical implications – On housing prices emanating from proximity to segregated areas with high income differs from segregated areas with low income, policies that address socio-economic costs and benefits, as well as property assessment levels, should reflect this pronounced difference. On the property level, positive spillover on housing prices near high-income segregated areas will cause an increase in the number of higher income groups and exacerbate segregation based on income. Contrarily, negative spillover on housing prices near low-income areas might discourage high-income households from moving to a location near low-income segregated areas. Local government should be aware of these spillover effects on housing prices to ensure that policies intended to reduce socioeconomic segregation, such as residential and income segregation, produce desirable results.

Social implications – Furthermore, a good estimation of these spillover effects on housing prices would allow local governments to carry out a cost–benefit analysis for policies intended to combat segregation and invest in deprived communities.

Originality/value – The main contribution of this paper is to go beyond the traditional studies of segregation that mainly emphasise residential segregation based on income levels, i.e. low-income or high-income households. The authors have analysed the spillover effect of proximity to hot spots (high income) and cold spots (low income) on the housing values of nearby condominiums or single-family homes within segregated areas in Stockholm Municipality in 2013.

Keywords Segregation, Spillover, Housing values

Paper type Research paper

1. Introduction

Segregation refers to the residential separation of population groups based on characteristics such as ethnicity, race, socioeconomic status, age and income. Although a wide range of population groups can be subject to segregation, research and literature on segregation have focused on population groups whose spatial segregation causes political and social problems (Fujita and Maloutas, 2016).

The focus of segregation studies in the USA, where segregation remains a major concern, has been on the segregation of racial and ethnic groups. This focus is related to the conditions of American cities at the beginning of the 20th century, where there is the legacy of slavery and discriminatory practices against racial and ethnic groups, including residential segregation. Although this situation gradually changed after Second World War, the established patterns of separation along racial-ethnic lines did not change. However, in European cities, which are more homogeneous in racial and ethnic terms, studies in the first decades after the Second World War focused on social class as an identifier of spatial separation, especially with labour migration (Fujita and Maloutas, 2016).

Based on country-of-origin statistics, provided by Statistics Sweden, segregation has increased over time. On average, the segregation index is equal to 23 in 2017, which means that 23% of the population would need to relocate for even distribution (segregation index equal to 0). To create the index, the municipality is divided into smaller geographical areas. Distribution of native-born and foreign-born persons and the deviation area by area is calculated and compared to the municipality's average [1]. We can further note that the variation between municipalities is significant. The maximum and minimum segregation index in Sweden's municipalities is 46.8 and 1.6, respectively (Statistics Sweden (SCB), 2017). Part of this variation is undoubtedly explained by variation in income, the proportion of migrants, the level of housing prices and general access to housing.

Since the end of the Second World War, Swedish housing policy has focused on providing good and affordable housing for a large portion of the population. To achieve that, municipal housing companies were established. The public housing built by these municipal housing companies represents more than 20% of the total housing stock in Sweden. The most important initiative in public housing was the "Million Dwellings Programme". The purpose of this program was to build one million new homes in a 10-year period, between 1965 and 1974, ensuring that everyone could have a home at a reasonable price (defined as not more than one-quarter of the family's disposable income). However, a consequence was the increased concentration of more marginal groups, especially those with low incomes, in the suburban periphery. In the Stockholm region, housing was predominantly built in the suburbs during this period, with the northern and southern suburbs having grown significantly (Murdie and Borgegård, 1998).

The spatial distribution of different tenure types, whether concentrated in different neighbourhoods or mixed, determines the extent to which the housing market segmentation affects segregation (Andersen *et al.*, 2016) . Residential segregation by income has spillover

effects on housing values. O'Sullivan (2019) describes bidding behaviour of households for favourable neighbours and their preference to live neighbourhoods with large numbers of high-income and educated households. In neighbourhoods populated by high-income households, municipalities can collect higher taxes. Therefore, the schools and services improve, while social standards rise; this leads to even higher housing prices, and as a result, income segregation continues to increase (Fogli, and Guerrieri, 2019).

The main research question we seek to answer is whether there is a relationship between housing segregation and neighbouring housing prices – that is, a spillover effect. In other words, does proximity to a specific type of segregated housing market have a negative impact on nearby housing markets, while proximity to another type of segregated market has a positive impact? Analysis of spillover effects from segregated residential areas is rare. Some earlier studies (i.e. Lee *et al.*, 1999; Lyons and Loveridge, 1993) analyse the question and found out that affordable housing developments have a negative impact on property values, but the question has not been empirically analysed with the application of data from within Sweden. There are possible policy implications associated with any observed spillover effect, i.e. on urban planning and local government policies, social cohesion policies and the socioeconomic cost for those living in the segregated and nearby areas.

To answer this research question, we intend to analyse the relationship between housing segregation and its spillover effect on neighbouring housing prices in Metropolitan Stockholm. First, we estimate housing segregation based on income at a low, disaggregated level. We then analyse the causal effect on housing prices in neighbouring residential areas.

The main contribution of our paper is to extend the traditional studies of segregation, which mainly emphasise residential segregation based on income levels, i.e. low-income or high-income households. We have analysed the spillover effect of proximity to hot spots (high income) and cold spots (low income) on the housing values of nearby condominiums within segregated areas in Stockholm Municipality in 2014. One contribution is our research on different methods to estimate the causal connection. We not only estimate a hedonic price equation, but we also use the instrumental variable approach, spatial economics and treatment effect methodology with, among other things, propensity score (PS) methods.

The remainder of the paper is divided into five major sections. Section 2 provides a theoretical and conceptual framework for residential segregation and the spillover effect of housing segregation on neighbouring housing areas, with a review of previous studies. Section 3 explains our research methodology and different model approaches, as well as regression techniques used in estimating the relationship between housing segregation and housing values. Section 4 describes the different data sets we used, whereas Section 5 presents the empirical analysis of housing segregation by income and the possible causal effect on housing prices in neighbouring areas. The paper concludes with a discussion of the study's implications, the limitations of the study and possible routes for future studies.

2. Theoretical framework

Residential segregation between socioeconomic groups grew in Europe over the past several decades (Fujita and Maloutas, 2016; Musterd and Ostendorf, 1998; Tammaru *et al.*, 2016). Growing inequality in Europe presents a fundamental challenge and threatens the social sustainability of European urban communities and cities (Tammaru *et al.*, 2016).

Spatial segregation is also one of the most significant aspects that reflect the specific conditions of an urban housing market. Depending on the householders' preferences for a specific set of housing characteristics and demographic features, similar dwellings are concentrated spatially, and these preferences are bound by financial ability, where the home price/rent level dictates the housing options available to households. Thus, the dynamics of

the housing market reinforce the income-based focus of similar groups of households, which leads to spatial segregation based on income (Giffinger, 1998; Owens, 2019).

Income inequality represents the most critical and crucial cause of the residential segregation of socioeconomic groups (Musterd and Ostendorf, 1998; Quillian and Lagrange, 2016). The causal mechanism for income inequality includes three facets: first, a certain level of income inequality is generated by the labour market and ethnic–racial discrimination; second, income inequality and discrimination translate into social and racial segregation in the form of residential segregation; and third, residential segregation feeds back into inequality and discrimination (Fujita and Maloutas, 2016). Stockholm has seen a steep rise in the residential segregation of socioeconomic groups between the 1990s and 2000s (Tammaru *et al.*, 2016).

In addition to the socioeconomic factors, sociological explanations of other behavioural aspects to location decision-making can also explain certain housing markets dynamics and observed spatial segregation. The behavioural approach, the social and emotional perspective, have undeniable importance in decision-making in housing markets and influencing housing market dynamics (Schelling, 1971; Dunning, 2017; Salzman and Zwinkels, 2013). Schelling (1971) suggests that segregation can result from individual discriminatory behaviour that influences residential location decisions. The individual incentives, and perceptions of difference, can lead collectively to segregation, and thus individual disorganised behaviour can translate into collective outcomes. Bakens and Pryce (2019) and Savage (2010) studied how homophily affects homeowners' patterns of spatial relocation. Homophily horizons are probably crucial in creating the social structure of cities in the long term and driving segregation in the housing market (Bakens and Pryce, 2019).

A related issue analysed in the literature is the detrimental effect of proximity to affordable housing on housing values. Nguyen's (2005) literature review indicates that the presence of affordable housing alone is sufficient to impact local housing values. However, the effect depends mainly on how affordable housing is designed, the management of the buildings, the characteristics of the host neighbourhood and its compatibility with affordable housing, as well as the concentration of affordable housing. Similar results could be considered more generally for proximity to segregated areas. The effect of affordable housing, public housing, rental housing or segregated neighbourhoods can be characterised as a *not in my backyard* (NIMBY) phenomenon (see, e.g. Brunes *et al.*, 2020). Lyons and Loveridge's (1993) results indicated that locating affordable housing near residential property has a statistically significant negative effect on residential property values; however, this effect diminishes with greater distance and tends to be relatively small.

Proximity to affordable housing does have a strictly negative impact on property values. High-quality, affordable housing that is well-designed and well-maintained, as well as rehabilitated housing, has the potential to impact property values positively. Alternatively, proximity to this type of housing did not, at the very least, have any harmful effects on property values, nor did it cause property values to decrease, according to several studies (Briggs *et al.*, 1999; Goetz *et al.*, 1996; Santiago *et al.*, 2001).

3. Methodology

The primary purpose of this study is to estimate the relationship between housing values and proximity to segregated housing areas, i.e. the presence of spillover effect. The model approach is the hedonic price model, where housing values are related to several characteristics of the home and the residential area. We are primarily interested in neighbourhood characteristics that are adjacent to the residential area. Our goal is to estimate the causal relationship. This is often problematic because of endogeneity problems,

which may themselves be caused by the lack of relevant attributes and simultaneity problems. We have tackled the problem of endogeneity in several different ways that will be briefly explained.

Of course, we must include all relevant attributes in the hedonic pricing model. In addition to these, we have also included fixed effects, to minimise the problem of omitted variable bias. We have also estimated different types of spatial models that aim to capture a spatial dependency. Furthermore, we used two different instrumental variable methods to address any endogeneity concern. A two-stage model has been estimated with, first, quasi-instrumental variables together with fixed effects, and second, with instrumental variables based on spatially lagged independent variables. Finally, we have estimated a model where we have reduced the sample, and only include observations in the vicinity of the segregated areas. It has similarities with a classic treatment-effect model. Here we will use the propensity score approach (PSA) to find similar comparison objects as often as possible.

3.1 Hedonic model

Like many previous studies, we will use the classic hedonic method to estimate how proximity to segregated areas affects housing prices. The hedonic method, developed and presented in a seminal article by Rosen (1974), has been used to estimate the impact of area characteristics such as traffic noise, proximity to parks and proximity to other types of positive or negative externalities. What we do, in principle, is to estimate a regression model where the home price is explained by several of the home's attributes, such as size, as well the residential area's attributes, such as its crime rate, or its proximity to retail and public transport. Here, the focus will be on the area attribute proximity to segregated housing areas. The model that will be estimated can be illustrated in the following equation:

$$P_i = \alpha + \beta_1 X_i + \beta_2 S_i + \beta_3 D_i + \varepsilon_i \qquad (1)$$

where P is equal to price, and X is a vector of housing and neighbourhood characteristics. S is the distance or proximity to the segregated residential area, and D is a vector of binary fixed effects. All letters in Greek are parameters to be estimated, and subscript i indicates the number of observations. In the first basic model, we are using the ordinary least square (OLS), and we are assuming, among other things, that all independent variables are exogenously given. This may not be the case. The choice of function form of the hedonic price equation is essential. While the theory says very little about the choice of function form, it can be considered as an empirical question. Here we have chosen to test the function form with a Box–Cox transformation (see, for example, Cropper *et al.*, 1988).

3.2 Spatial model

Transactions of home sales usually show a relatively high degree of spatial dependence. Although we have included fixed housing area variables in the model, there is a risk that we still have problems with spatial dependence (see, for example, Wilhelmsson, 2002, and Anselin, 2010). Here, like many other studies, we will test whether the residuals from the hedonic model show a spatial dependence with the help of Moran's I. If it turns out that we have problems with spatial dependence, two types of models will be estimated, namely, spatial error model (SEM) and the spatial autoregressive model (SARM). These models can be illustrated by equations (2) and (3), respectively:

$$P_i = \alpha + \beta_1 X_i + \beta_2 S_i + \beta_3 D_i + \rho W P_i + \varepsilon_i \tag{2}$$

$$P_i = \alpha + \beta_1 X_i + \beta_2 S_i + \beta_3 D_i + \lambda W \varepsilon_i + u_i \tag{3}$$

where W is a spatial weight matrix. Here we are using the inverse distance between the observations as a weight matrix. The parameter ρ is the coefficient of the spatially lagged dependent variable, and measures the spatial dependency between observations i and j. The parameter λ is the coefficient in a spatial autoregressive structure for the disturbance ε. The parameters will be estimated using the maximum likelihood method.

3.3 Instrumental variable approach

In the event that one or more independent variables are not exogenously given, the instrument variable (IV) method has been highlighted as a potential solution to this problem. The idea is to find variables (instruments) that are correlated with the exogenous variable but not with the dependent variable. However, it has been proved notoriously difficult to find strong IV and, above all, there is no method of ascertaining whether selected instruments are strong or weak. See Bascle (2008) for a review of controlling endogeneity with instrumental variables.

In our study, it has also been challenging to find suitable IV. What we have done is to use something called quasi-instruments (Bartels, 1991), and IVs that are spatially lagged variables of the independent variables included in the hedonic equation. This IV approach can be divided into two steps. First, we estimate a model where the independent variable of interest is related to the instrumental variables and all other variables. Second, the predicted value is used as an independent variable in the hedonic price equation. The estimated model can be illustrated as in equations (4) and (5):

$$P_i = \alpha + \beta_1 X_i + \beta_2 \hat{S}_i + \beta_3 D_i + \varepsilon_i \tag{4}$$

$$\hat{S}_i = \phi_0 + \phi_2 X_i + \phi_2 D_i + \phi_3 Z \tag{5}$$

where Z is a vector of instrumental variables.

3.4 Treatment effect model

To minimise the problem of endogeneity, we have reduced the sample to include only observations close to the segregated areas. This means that we consider those who are close to the segregated area as the treatment group, while those further away are the control group. Of course, this is not a difference-in-difference approach we use, but a type of treatment-effect model, where transactions in the treatment group have the value one, and in the control group, the value is zero. In the first step, we have estimated models using different distances for the treatment and control groups. Here we have chosen distances of 500, 750 and 1,000 m for the treatment group, and of 1,500 and 2,000 m for the control group. As a selection variable regarding which assumption is most optimal, we have chosen the AIC value.

We have chosen to use the PSA in making the comparison between the treatment group and the control group as similar as possible, alongside the treatment variable itself. This is a method proposed by Rosenbaum and Rubin (1983) to mitigate bias in the estimation of treatment effects with non-randomised data. See Austin (2011) for a review of different PSAs. We have chosen to use it in two ways: in a multivariate approach where we include

PSs directly in the hedonic regression model, and when we use the inverse of the PS as sample weights (see, for example, Wilhelmsson, 2019; Hyland *et al.*, 2013; Reichardt, 2013; Hirano *et al.*, 2003). Moreover, we used the technique of PS matching, where we used the nearest neighbour (NN) matching. Only matched observations are used in the hedonic price equation (see, for example, Wilhelmsson, 2019; Holupka and Newman, 2012). The model has been estimated in two steps, wherein the first was calculating the probability that the observation has received treatment with logistic regression. This probability is then used in the hedonic equation directly as an independent variable, or as sample weights. Estimated models are illustrated as in equations (6) and (7):

$$P_i = \alpha + \beta_1 X_i + \beta_2 Treat + \beta_2 D_i + \beta_i Prob_i + \varepsilon_i \tag{6}$$

$$Prob(Treat|X)_i = \eta_0 + \eta_1 X_1 + \eta_2 D_i \tag{7}$$

where *Treat* is equal to 1 if the dwelling is in close proximity to the segregated area.

3.5 Segregation variable
Of central interest is the variable that measures segregation or rather the "hot spot" of segregation. The hedonic price equation includes either the distance to a segregated area, or a binary variable that indicates that the observation is located adjacent to a segregated area. We are using Getis-Ord statistics G_i^* as a measure of a segregated residential area.

Getis-Ord G_i^* statistics have been used to study and analyse evidence of identifiable spatial patterns. G_i^* index is similar to other indices such as Moran's I and Geary's C. However, Moran indices do not distinguish between hot spots and cold spots. G_i^* index can determine cluster structures of hot and cold spots concentration among local observations. In other words, given a set of weighted features, G_i^* identifies statistically significant hot spots and cold spots. G_i^* is usually standardised based on its sample mean and variance:

$$Z\left(G_i^*\right) = \frac{\sum_{j=1}^n w_{i,j} x_j - \overline{X} \sum_{j=1}^n w_{i,j}}{S\sqrt{\frac{n\sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j}\right)^2}{n-1}}} \tag{8}$$

Consider a study area subdivided into n regions $i = 1,2,\ldots,n$, where each region is identified with known Cartesian coordinates with measurements $X = [x_1,\ldots,x_n]$. Each i value has an associated x_i value. Moreover, the weights w_{ij} are defined between all pairs of points i and j (for all $i, j \in \{1,\ldots,n\}$). The standardised G_i^* statistic for each feature is basically a Z score and, therefore, can be attached to the statistical significance. A high positive Z score for a feature indicates a spatial clustering of (hot spots). A low negative Z score indicates a spatial clustering of (cold spots), whereas a Z score near zero indicates no apparent spatial clustering (Ord and Getis, 1995; Songchitruksa and Zeng, 2010).

The distance to hot/cold spots has been calculated as the smallest Euclidean distance between a property and a hot/cold spot where the hot/cold spots are defined according to equation (8) in all 250×250 m squares (see data description in Section 4).

4. Data and descriptive statistics
For the purposes of this paper, we must combine information on segregation within a city with information on property values in that city. Therefore, we have used income data

together with data on housing values taken from residential property transactions. The case study used is the Municipality of Stockholm, the capital of Sweden. The following sections present the two data sets with descriptive statistics and illustrate how we estimated the concentration of segregated areas.

4.1 Disaggregated income data and the measurement of segregation

To analyse housing segregation, we have used the information about the income of residents of Stockholm Municipality presented in Table 1. The data is spatial to the extent that it provides information at a disaggregated level for squares measuring 250 × 250 m. The total number of squares is 2,421, with information on the population and their incomes.

As Table 1 demonstrates, the observed increases are distinct among various income groups and over time. From 2000 to 2013, the average number of people with high and upper middle income increased 16% and 13%, respectively, whereas the average number of people with low and lower middle income increased 11% and 9%, respectively. From 2000 to 2007, the average number of low- and high-income individuals in the included area increased by 5% and 4%, respectively. There were no significant changes in the average number of people with lower middle income and upper middle income during the same period. From 2007 to 2013, the observed 23% increase of the average number of upper middle-income and high-income people was higher (12% and 11%, respectively), whereas the increase in the average number of people with low income and lower middle income was 6% and 8% respectively.

The data encompasses most of the past two decades, which allows us to analyse trends over time. Our methods are partly to calculate different types of segregation measures and analyse them over time or in space, using the Index of Dissimilarity and the Information Theory Index, as well as the Getis-Ord statistics.

Table 2 shows income segregation in Stockholm between 2000 and 2013. Income segregation measured by, e.g. the Index of Dissimilarity and the Information Theory Index, has increased over time. These measures are the most widely used methods to measure residential segregation. The Index of Dissimilarity measures the extent to which the population within each group is distributed evenly across the region. It is also used to measure inequality, which is known as relative mean deviation. The Information Theory Index is used to analyse segregation within and between communities, and as a measure of

| Year | 2000 | | 2007 | | 2013 | |
Variables	Mean	SD	Mean	SD	Mean	SD
Total number of people with income	280.645	343.412	288.337	346.530	315.570	369.480
Number of people with low income in the included area	70.189	97.078	73.779	101.308	77.965	105.152
Number of people with lower middle income in the included area	56.188	68.472	56.240	67.594	60.972	71.502
Number of people with upper middle income in the included area	58.856	69.556	59.378	69.340	66.345	75.566
Number of people with high income in the included area	95.412	132.904	98.940	141.137	110.287	155.319
N	2,421		2,421		2,421	

Table 1.
Population of the municipality of Stockholm in 2000, 2007 and 2013

Notes: Income levels, in increments of one thousand Swedish krona, are defined as follows: low income (<150), lower middle (150–250), upper middle (250–360) and high income (>360). Income of individuals below 20 years of age was not considered

segregation, the index measures the diversity of local areas compared to the overall diversity of a region, instead of measuring the diversity between the local and overall proportions of each group. Index values typically range between 0 and 1 (Roberto, 2015).

For the Municipality of Stockholm, in 2013, we can observe in Figure 1 the substantial concentration and size of the measurements made with the Getis-Ord statistics (hot and cold spots) in the segregated areas. Areas marked in blue denote a more significant proportion of lower income households (cold spots), whereas red denotes a more significant proportion of high-income households (hot spots).

4.2 Transaction data and the measurement of segregation spillover
The empirical analysis will be the estimation of the traditional hedonic pricing model. It will be estimated for the condominium market. Data refer to the year 2014 in the Municipality of

Year	Dissimilarity index	Information theory index
2000	0.1777	0.0457
2007	0.2003	0.0568
2013	0.2047	0.0582

Notes: Dissimilarity index = a measure of how different the social composition of a neighbourhood is, on average, from the social composition of the study area, where 0 is no segregation. Information theory index = a measure of how much less socially diverse a neighbourhood is, on average, than the study area, where 0 is no segregation

Table 2.
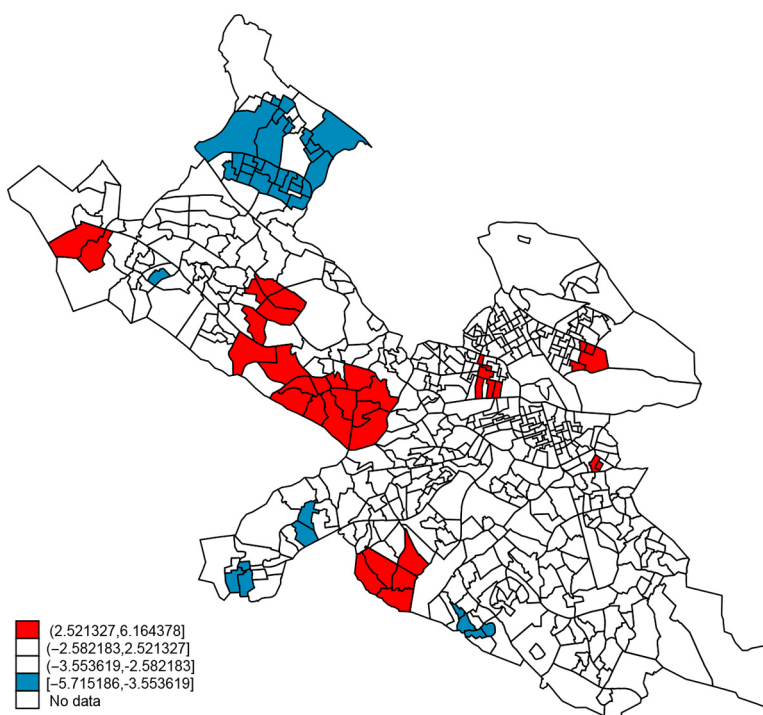Diversity measures:
income

Figure 1.
Hot (high income) and
cold (low income)
spot segregation in
Stockholm
Municipality, 2013

Stockholm. The source of these data is Mäklarstatistik AB [2] (Swedish Brokerage Statistics), an important source of statistical information for the Swedish housing market. Since 2005, Mäklarstatistik has been collecting data from home sales that take place via brokers. Almost all brokers in Sweden report their contract data to Mäklarstatistik, which enables Mäklarstatistik to present highly up-to-date and comprehensive price statistics for homes throughout Sweden. This provides an overall picture of price development in the Swedish housing market. Descriptive statistics regarding the included variables in the hedonic model are illustrated in Table 3.

Table 3 presents the descriptive statistics regarding the data on condominium sales. The total number of observations is 12,621. The average price is just over SEK 3m, with a standard deviation of as much as SEK 1.7m. The average size of rental apartments is 62 square meters, with a standard deviation of 24 square meters. As the owner of a condominium, you pay a monthly fee to the housing association to cover the cost of common areas and ongoing maintenance. On average, this fee amounts to SEK 3,200 with a variation corresponding to one-third of that amount. We have included distance to 39 different shopping malls classified as city malls, outlet malls, regional malls, regional retail parks, super-regional malls or theme centres by the Swedish Shopping Center Directory, provided by the firm Datscha (see also Long and Wilhelmsson, 2020). Distance to the nearest metro station is also included in the hedonic model. The average distance to the shopping mall and metro station was 1,430 and 550 m, respectively. Most observations are close to the central business district (CBD), with an average distance of 4.7 km. The average distance to a segregated residential area, whether one with low or high incomes, amounts to about 1,500 m; however, the variation is significant.

5. Econometric analysis
The econometric analysis will aim to estimate the hedonic equation. We will use the Box–Cox transformation to find the empirically best form of the hedonic function. The proximity to hot and cold spots will be included as a continuous variable in a hedonic price model and as a binary variable in a so-called treatment effect model.

5.1 Hedonic price equation
Here, we will estimate the hedonic price model. Given the results of the Box–Cox analysis, we cannot reject a natural logarithm transformation of the price, but we reject the same transformation of the independent variables. Hence, a semi-logarithmic specification form of

Variables	Mean	SD	Min	Max
Price (SEK)	3,357 363	1,738,447	300,000	16,500,000
Living area (square metres)	61.89	23.74	22	147
No. of rooms (no.)	2.37	0.96	1	7
Monthly fees (SEK)	3,218.86	1,232.79	866	7,247
Distance to CBD (km)	4.662	3.145	0.293	15.218
Distance to hot spot (m)	1,482.04	1,108.68	2.85	4,166
Distance to cold spot (m)	1,311.86	821.44	9.9	2,993.12
Distance to shopping mall (m)	1,430	950	20	4,460
Distance to metro (m)	550	400	3	2,120

Table 3.
Descriptive statistics

Notes: The number of observations is 12,621. Prices are given in Swedish krona (SEK), and distance is measured in metres

the hedonic pricing model is preferred, and all models in this analysis use a semi-logarithmic form of the hedonic pricing equation. Table 4 shows the results of the default OLS model, as well as for the two instrumental variable models (IV1 and IV2) and the two spatial models (SAR and SEM).

In the models, we can note that the goodness-of-fit is high at around 87%. Moran's I statistics show spatial dependency in the OLS model, even if we control for spatial location by including both zip codes and coordinates in the model (not shown in the Table). All estimated parameters have expected sign and reasonable magnitude. For example, increasing the living area by one square meter increases the price by about 1.4%. The further away from the CBD the apartment is located, the lower the price. The same negative effect applies for distance to shopping mall and metro stations.

The magnitude of the effect is very robust across models. That is to say, the instrumental variable approach does not, in any significant way, produce effects that are higher or lower compared to the OLS model. However, the expected effect of distance to the shopping mall is almost double in magnitude in the spatial models when compared to the OLS and the IV models. Almost all the estimated parameters are statistically significant.

Proximity to "hot" and "cold" spots is included as a continuous variable in the hedonic pricing equation. A hot spot is defined as a segregated high-income area, and cold spot as a segregated low-income area. When it comes to distances to segregated areas, we can see a negative spillover effect from a segregated area with low incomes. The parameter estimate is statistically significant in all models. The spillover effect from high-income areas is statistically significant in all models, but the impact switches sign. In the two spatial models, it can be observed that proximity to the hot spot has a positive effect insofar as the parameter estimate is positive, i.e. the further away from a hot spot, the higher the apartment's value, all other things being equal. The estimates regarding proximity to a cold spot are all statistically significant, but the effect differs depending on which estimation method we use. The effect is equivalent in the OLS, IV1 and SEM models, whereas it is higher in the IV2 model and lower in the SAR model. The conclusion is that proximity to a low-income segregated area has a negative effect, i.e. the further away from, the higher the apartment's value, all other things being equal. However, the effect is not robust.

5.2 Treatment effect model

In models presented, the goal has been to minimise the endogeneity problem by estimating instrument variable models and spatial models. In a further attempt to reduce the problem of endogeneity, we have also estimated so-called treatment effect models. Hence, the relationship between segregation and housing values is analysed by creating a binary treatment variable, where 1 signifies that the house is close to a segregated area, and 0 that the house is elsewhere. The control group thus consists of housing that is further away from the segregated areas, where no spillover effect is expected. This is done to reduce the problem of spatial dependence and omitted variable bias, i.e. to reduce the problem of endogeneity ultimately.

Several distance measures concerning the treatment and control areas will be tested empirically. The treatment group will consist of housing 500, 750 and 1,000 m from the segregated area, and the control group will be 1,500 and 2,000 m from the segregated area, respectively. Through a grid search, we have chosen the one that minimises the AIC value, which is the specification where treatment is in the range of 125–1,000 m, and the control range is 1,000–2,000 m.

When we measure the spillover effect, there is a risk that some housing is located in the segregated area, and that is not desirable. We have, therefore, included a buffer zone around the segregated area. Any housing located within 125 m of the segregated area may be

	(1) OLS	(2) IV1	(3) IV2	(4) SAR	(5) SEM
Living area	0.0137*** (89.08)	0.0138*** (88.60)	0.0137*** (85.72)	0.0145*** (89.37)	0.0133*** (87.01)
Monthly fee	−0.0000523*** (−24.30)	−0.0000553*** (−25.38)	−0.0000549*** (−24.57)	−0.0000655*** (−29.56)	−0.0000539*** (−23.64)
Rooms	0.0267*** (7.97)	0.0272*** (8.02)	0.0283*** (8.13)	0.0210*** (5.80)	0.0340*** (10.66)
CBD	−0.0578*** (−19.25)	−0.00185*** (−8.42)	−0.00364*** (−8.11)	−0.0809*** (−80.75)	−0.0859*** (−83.61)
Shopping mall	−0.0233*** (−6.94)	−0.0286*** (−8.27)	−0.0429*** (−9.15)	−0.0531*** (−24.32)	−0.0296*** (−8.84)
Metro station	−0.0221*** (−4.29)	−0.0289*** (−5.37)	0.00366(0.42)	−0.0478*** (−11.68)	−0.0250*** (−4.20)
Hot spot	0.0000162*** (3.48)	−0.0000197* (−2.44)	−0.0000181(−1.89)	−0.00000476* (−2.26)	0.0000166*** (4.97)
Cold spot	0.0000683*** (15.61)	0.0000557*** (7.19)	0.000131*** (6.77)	0.0000268*** (10.18)	0.0000586*** (13.59)
Constant	21.98 (1.36)	−2.368 (−0.15)	−21.24(−1.15)	14.62*** (832.36)	14.54*** (1134.11)
ρ				0.00133 (1.81)	
λ					3.079*** (863.26)
Observations	12,621	12,621	12,621	12,621	12,621
R^2	0.881	0.878	0.872	0.8577	0.8535
AIC	−10,996.4	−10,668.5	.	−8,800.2	−11,541.5
Moran's I	2,175.91				

Notes: t statistics in parentheses; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 4.
Regression results, OLS and IV-estimates, SAR and SEM

considered to be within the area. Thus, we will only include housing that is at least 125 m from a segregated area. However, this means that some housing located near but outside the segregated area will not be included in the analysis. This fact means that we are introducing bias, but this is expected to be less than the bias that would occur if we include all housing less than 125 m from a segregated area in our model. We have also separated the estimates of proximity into one segregated area with high incomes and one area with low incomes. The results from these models are presented in Tables 5 and 6.

In the treatment model (TE) presented in Table 5, we can see that the degree of explanation is remarkably high. About 88% of the variation in price can be explained by the variables used. Compared to hedonic models, significantly fewer observations are used in the calculation. Instead of over 12,000 observations, we are here using only 6,800. As mentioned earlier, this is to reduce the risk of omitted variable bias and spatial dependency. All coefficients have the expected signs and are comparable to the earlier models. The estimate for treatment is negative and statistically significant. The parameter estimates amount to 0.0925, which corresponds to 9.2% of the value of the apartment.

Table 5 also shows two models that use the PS methodology. In the first step, the probability of being located in the treatment area has been calculated using logistic regression. This means that we have tried to identify objects comparable to those in the treatment area. In the first model (PS1), these PSs are included as an independent variable in the model, and in the second (PS2), the model has been calculated with PSs as sample weights. The degree of explanation is slightly higher in these models compared to the TE model. The results from both models prove to be robust with the TE model. In the last column of Table 5, the results from a model that uses the NN as a comparison object in the control area are reported. This means that we have matched each observation in the treatment area with an observation in the control area that is as similar as possible, regardless of whether they have treatment or not. The number of observations used in this model is only 2,494. The result from this model (NN) turns out to have a slightly lower degree of explanation, compared with previous models, and the value estimate regarding treatment is also slightly less, approximately 8.5% lower in the treatment area compared with the control area.

Table 6 shows the results from similar models, namely, TE, PS1, PS2 and NN, but here we have instead calculated the effect of being located near a hot spot. In general, the degree of explanation is high in all models, around 88%. The estimates regarding the independent variables vary more, compared to the model estimating the effect of cold spots. These estimates are not as robust as when we analysed cold spots. However, all estimates have the expected signs and are statistically significant. The estimate regarding treatment is, as expected, positive, which means that apartments that are close to, but not within, high-income areas, have a higher value compared to apartments further away. The estimates are relatively robust between the different models, and the house prices are approximately 2% higher in the treatment area compared to the control area. The only model that shows non-significant estimates regarding treatment is the NN model.

In summary, we can conclude that there is a spillover effect on nearby housing values from segregated housing areas. In segregated areas with low incomes, this spillover effect is negative, whereas for areas with high incomes, the spillover effect is positive.

6. Conclusion and policy implications
The purpose of the present study was to identify residential areas where the segregation of low-income or high-income households occurs. We used Getis-Ord statistics to identify the so-called hot and cold spots. By identifying areas of the city where there is a concentration of households with similar incomes, it was possible to investigate whether proximity to these areas spills over into the housing values of nearby condominiums.

	(1) TE	(2) PS1	(3) PS2	(4) NN
Living area	0.0139*** (68.28)	0.0135*** (59.39)	0.0128*** (52.35)	0.0137*** (32.78)
Monthly fee	−0.0000532*** (−18.60)	−0.0000421*** (−11.27)	−0.0000322*** (−8.17)	−0.000704*** (−12.92)
Rooms	0.0220*** (4.99)	0.0252*** (5.68)	0.0489*** (8.79)	0.0108(1.32)
CBD	−0.0536*** (−13.26)	−0.0492*** (−11.87)	−0.0590*** (−12.88)	−0.0400*** (−5.08)
Shopping mall	−0.0389*** (−8.41)	−0.0548*** (−9.53)	−0.0211*** (−3.50)	−0.0135(−1.52)
Metro station	−0.0264*** (−3.88)	0.0365* (2.41)	−0.0260* (−2.46)	−0.0751*** (−6.28)
Treatment	−0.0926*** (−15.53)	−0.0915*** (−15.35)	−0.0899*** (−8.28)	−0.0857*** (−7.20)
Propensity score		−0.255*** (−4.64)		
Constant	−209.0*** (−9.03)	−83.35* (−2.34)	−163.0*** (−5.05)	−354.5*** (−8.91)
Observations	6,846	6,846	6,846	2,494
R^2	0.882	0.882	0.917	0.809
AIC	−6,732.4	−6,752.0	−8,576.2	−1,907.1

Notes: *t* statistics in parentheses; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

Table 5.
Treatment effect
models (treatment =
close to cold spots)

	(1) TE	(2) PS1	(3) PS2	(4) NN
Living area	0.0140*** (67.38)	0.0113*** (31.09)	0.0137*** (48.90)	0.0142*** (51.48)
Monthly fee	-0.0000537*** (-18.47)	-0.0000795*** (-19.68)	-0.0000635*** (-14.15)	-0.000503*** (-13.13)
Rooms	0.0213*** (4.76)	0.0637*** (9.91)	0.0261*** (4.10)	0.0265*** (4.56)
CBD	-0.0422*** (-10.46)	-0.0654*** (-13.78)	-0.0397*** (-8.36)	-0.0622*** (-9.71)
Shopping mall	-0.0214*** (-4.62)	-0.164*** (-10.08)	-0.0429*** (-7.13)	0.000656 (0.10)
Metro station	-0.0514*** (-7.61)	-0.0922*** (-11.44)	-0.00274 (-0.26)	-0.0543*** (-5.60)
Treatment	0.0202*** (4.24)	0.0321*** (6.52)	0.0256*** (4.76)	0.000165 (0.02)
Propensity score		-1.210*** (-9.14)		
Constant	-249.4*** (-10.67)	102.4* (2.28)	-238.6*** (-8.09)	-234.5*** (-7.47)
Observations	6,846	6,846	6,846	3,777
R^2	0.878	0.879	0.883	0.880
AIC	-6,512.5	-6,593.9	-6,320.1	-4,027.2

Notes: t statistics in parentheses; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 6.
Treatment effect
models (treatment =
close to hot spots)

Why is it of interest to investigate this issue, and what policy implications does it have? Regardless of whether segregation is in a state of stable equilibrium, segregation brings socioeconomic costs, not only for those living in the segregated residential areas, but also for nearby areas, and the city as a whole. The spillover effect is one such socioeconomic cost and is, therefore, an essential aspect of segregation to measure and analyse. Of course, segregated areas are not just areas with lower income households; there are, of course, segregated high-income areas. We have chosen to identify segregated areas with both high and low incomes in our case study examining the Municipality of Stockholm, Sweden.

In the period we studied (2000 through 2013), we can see that segregation has increased. With the help of Getis-Ord Statistics, we have identified hot and cold spots in Stockholm for 2013. We have then matched these segregated areas with sales data for condominiums. This has made it possible to calculate the minimum distance from each transaction to any segregated residential area with high and low income. We have then estimated a traditional hedonic pricing equation. Whether we estimate using a hedonic model or a so-called treatment effect model, we can observe an apparent positive effect for proximity to segregated areas with high income, and a negative capitalisation effect in the vicinity of segregated areas with low incomes.

The policy implications of the findings of this research are manifold. Because the identified spillover effect on housing prices emanating from proximity to segregated areas with high income differs from segregated areas with low income, policies that address socio-economic costs and benefits, as well as property assessment levels, should reflect these pronounced differences.

On the property level, positive spillover on housing prices near high-income segregated areas will cause an increase in the number of higher income groups and exacerbate segregation based on income. Contrarily, negative spillover on housing prices near low-income areas might discourage high-income households from moving to a location near low-income segregated areas. Local government should be aware of these spillover effects on housing prices to ensure that policies intended to reduce socioeconomic segregation, such as residential and income segregation, produce desirable results.

A better understanding of the different potential spillover effects on housing prices in relation to the spatial distribution of various income groups would be beneficial in determining appropriate property assessment levels. In other words, awareness of this spillover effect could improve existing property assessment methods, and provide local governments with extra information to make informed decisions about policies and services needed in different neighbourhoods.

Furthermore, a good estimation of these spillover effects on housing prices would allow local governments to carry out a cost–benefit analysis for policies intended to combat segregation and invest in deprived communities.

Because of distinctions between condominiums and single-family houses, it is plausible that any spillover effect emanating from proximity to segregated areas will be different for the two types of properties. Thus, further research that focuses on the impact of possible spillover on single-family house prices would be a valuable contribution to this research topic. Other important concerns regarding the spillover effect on house prices, which need to be investigated, are its impact on the future viability of development, and the way that the spillover effect influences the planning system as well as subsidies for apartment buildings.

Notes

1. www.scb.se/hitta-statistik/temaomraden/statistik-om-integration/regeringsuppdraget-registerdata-for-integration/sa-har-ar-statistiken-om-integration-uppbyggd/om-statistiken---integration-och-boende/

2. https://www.maklarstatistik.se

References

Anselin, L. (2010), "Thirty years of spatial econometrics", *Papers in Regional Science*, Vol. 89 No. 1, pp. 3-25.

Austin, P.C. (2011), "An introduction to propensity score methods for reducing the effects of confounding in observational studies", *Multivariate Behavioral Research*, Vol. 46 No. 3, pp. 399-424.

Bakens, J. and Pryce, G. (2019), "Homophily horizons and ethnic mover flows among homeowners in Scotland", *Housing Studies*, Vol. 34 No. 6, pp. 925-945.

Bartels, L.M. (1991), "Instrumental and 'quasi-instrumental' variables", *American Journal of Political Science*, Vol. 35 No. 3, pp. 777-800.

Bascle, G. (2008), "Controlling for endogeneity with instrumental variables in strategic management research", *Strategic Organization*, Vol. 6 No. 3, pp. 285-327.

Briggs, X., Darden, J.T. and Aidala, A. (1999), "In the wake of desegregation: early impacts of scattered-site public housing on neighborhoods in Yonkers, New York", *Journal of the American Planning Association*, Vol. 65 No. 1, pp. 27-49.

Brunes, F., Hermansson, C., Song, H.S. and Wilhelmsson, M. (2020), "NIMBYs for the rich and YIMBYs for the poor: analysing the property price effects of infill development", *Journal of European Real Estate Research*, Vol. 13 No. 1, pp. 55-81.

Cropper, M.L., Deck, L.B. and McConnell, K.E. (1988), "On the choice of functional form for hedonic price functions", *The Review of Economics and Statistics*, Vol. 70 No. 4.

Dunning, R.J. (2017), "Competing notions of search for home: behavioural economics and housing markets", *Housing, Theory and Society*, Vol. 34 No. 1, pp. 21-37.

Fogli, A. and Guerrieri, V. (2019), "The end of the American dream? Inequality and segregation in us cities (no. w26143)", *National Bureau of Economic Research*.

Fujita, K. and Maloutas, T. (2016), *Residential Segregation in Comparative Perspective: Making Sense of Contextual Diversity*, Routledge, London.

Giffinger, R. (1998), "Segregation in vienna: impacts of market barriers and rent regulations", *Urban Studies*, Vol. 35 No. 10, pp. 1791-1812.

Goetz, E.G., Lam, H.K., and Heitlinger, A. (1996), *There Goes the Neighborhood? The Impact of Subsidised Multi-Family Housing on Urban Neighborhoods*, Center for Urban and Regional Affairs and Neighborhood Planning for Community Revitalization, University of MN, Minneapolis, MN.

Hirano, K., Imbens, G.W. and Ridder, G. (2003), "Efficient estimation of average treatment effects using the estimated propensity score", *Econometrica*, Vol. 71 No. 4, pp. 1161-1189.

Holupka, S. and Newman, S.J. (2012), "The effects of homeownership on children's outcomes: real effects or self-selection?", *Real Estate Economics*, Vol. 40 No. 3, pp. 566-602.

Hyland, M., Lyons, R.C. and Lyons, S. (2013), "The value of domestic building energy efficiency-evidence from Ireland", *Energy Economics*, Vol. 40, pp. 943-952.

Lee, C.-M., Culhane, P. and Wachter, S. (1999), "The differential impacts of federally assisted housing programs on nearby property values: a Philadelphia case study", *Housing Policy Debate*, Vol. 10 No. 1, pp. 75-93.

Long, R. and Wilhelmsson, M. (2020), "Impact of shopping malls on apartment prices: the case of Stockholm", *Nordic Journal of Surveying and Real Estate Research Special Research*, Vol. 5, pp. 29-48.

Lyons, R.F. and Loveridge, S. (1993), "A hedonic estimation of the effect of federally subsidised housing on nearby residential property values", Staff Paper P93–6, University of Minnesota, Department of Agriculture and Applied Economics.

Murdie, R.A. and Borgegård, L.-E. (1998), "Immigration, spatial segregation, and housing segmentation of immigrants in metropolitan Stockholm, 1960-95", *Urban Studies*, Vol. 35 No. 10, pp. 1869-1888.

Musterd, S. and Ostendorf, W. (1998), *Urban Segregation and the Welfare State: Inequality and Exclusion in Western Cities*, Routledge, London.

Nguyen, M.T. (2005), "Does affordable housing detrimentally affect property values? A review of the literature", *Journal of Planning Literature*, Vol. 20 No. 1, pp. 15-26.

Ord, J.K. and Getis, A. (1995), "Local spatial autocorrelation statistics: distributional issues and an application", *Geographical Analysis*, Vol. 27 No. 4, pp. 286-306.

O'Sullivan, A. (2019), *Urban Economics*, 9th ed., McGraw-Hill Education, New York, NY.

Owens, A. (2019), "Building inequality: housing segregation and income segregation", *Sociological Science*, Vol. 6, p. 497.

Quillian, L. and Lagrange, H. (2016), "Socioeconomic segregation in large cities in France and the United States", *Demography*, Vol. 53 No. 4, pp. 1051-1084.

Reichardt, A. (2013), "Operating expenses and the rent premium of energy star and LEED certified buildings in the Central and Eastern US", *The Journal of Real Estate Finance and Economics*, Vol. 49 No. 3, pp. 413-433.

Roberto, E. (2015), "The divergence index: a decomposable measure of segregation and inequality", arXiv preprint arXiv:1508.01167.

Rosen, S. (1974), "Hedonic prices and implicit markets: product differentiation in pure competition", *Journal of Political Economy*, Vol. 82 No. 1, pp. 34-55.

Rosenbaum, P.R. and Rubin, D.B. (1983), "The central role of the propensity score on observational studies for causal effects", *Biometrika*, Vol. 70 No. 1, pp. 41-55.

Salzman, D. and Zwinkels, R.C.J. (2013), "Behavioral real estate", Doctoral thesis, Tinbergen Institute, available at: http://papers.tinbergen.nl/13088.pdf (accessed 15 February 2021).

Santiago, A.M., Galster, G.C. and Tatian, P. (2001), "Assessing the property value impacts of the dispersed subsidy housing program in Denver", *Journal of Policy Analysis and Management*, Vol. 20 No. 1, pp. 65-88.

Savage, M. (2010), "The politics of elective belonging", *Housing, Theory and Society*, Vol. 27 No. 2, pp. 115-161.

Schelling, T.C. (1971), "Dynamic models of segregation", *The Journal of Mathematical Sociology*, Vol. 1 No. 2, pp. 143-186.

Songchitruksa, P. and Zeng, X. (2010), "Getis–ord spatial statistics to identify hot spots by using incident management data", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2165 No. 1, pp. 42-51.

Statistics Sweden (SCB) (2017), available at: www.scb.se/en/

Tammaru, T., Marcińczak, S., van Ham, M., and Musterd, S. (2016), *Socioeconomic Segregation in European Capital Cities: East Meets West. Regional Studies Association*, Routledge, London.

Wilhelmsson, M. (2002), "Spatial models in real estate economics", *Housing, Theory and Society*, Vol. 19 No. 2, pp. 92-101.

Wilhelmsson, M. (2019), "Energy performance certificates and its capitalisation in housing values in Sweden", *Sustainability*, Vol. 11 No. 21, p. 6101.

Corresponding author
Mats Wilhelmsson can be contacted at: mats.wilhelmsson@abe.kth.se