Journal of Transport Geography 56 (2016) 102–109

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

# The impact of regional commuter trains on property values: Price segments and income

Helena Bohman [a,*,1], Désirée Nilsson [a,b,1]

[a] Department of Urban Studies, Malmo University, SE-205 06 Malmo, Sweden
[b] K2 The National Knowledge Centre for Public Transport, Scheelevägen 2, SE-223 81 Lund, Sweden

ARTICLE INFO

*Article history:*
Received 12 January 2016
Received in revised form 30 August 2016
Accepted 6 September 2016
Available online 16 September 2016

*Keywords:*
Hedonic price function
Price segment
Commuter railway
Two-stage quantile regression
Spatial lag

ABSTRACT

Using single-family home transactions and commuter rail data from 2014, we estimate hedonic price models using two-stage spatial quantile regression to capture variations across price segments. The results are significant and robust across different model specifications and across the different price segments, but the price effect of proximity to a commuter train station is strongest in lower price segments of the housing market. These price segment effects are also valid for proximity to highways, as well as for several other property attributes. Results also reveal that the largest of the three regional labour markets in our study has a greater effect on prices. Furthermore, the study introduces property-specific neighbourhood data from raster data, showing that population density has a negative impact on property prices at the neighbourhood level while population size has a positive impact at the municipal level.



## 1. Introduction

Urbanization pressures worldwide have resulted in dramatic increases in property values in many major cities, with a concomitant lack of affordable housing (UN-Habitat, 2016). Some of these pressures on urban property markets can be alleviated by improved accessibility, which gives those living on the outskirts of cities and in rural areas better access to labour markets, thus making these locations more attractive to households. Accessibility is a widely used concept, resulting in various definitions and alternative ways of operationalizing the concept (Rietveld and Bruinsma, 1998; Geurs and van Wee, 2004). However, it is generally agreed that improved accessibility is beneficial to the labour market as it improves efficiency through better matching of job requirements and individual skills. Households also benefit since they can opt for an urban work location with probably higher wages and choose a nonurban residence location where prices are lower (So et al., 2001). It is thus important to understand how property markets and different types of transportation interact, and thus affecting accessibility.

Studies of how railway accessibility affects property prices report mixed results. The studies do, however, differ in several respects, not least in the types of rail system studied, which range from high-speed trains (e.g. Andersson et al., 2010; Chen and Hall, 2011) to light rail or metro systems (e.g. Pagliara and Papa, 2011) and commuter rail systems (e.g. Debrezion et al., 2011). Moreover, there are differences in the type of land or property included in the study, data availability and quality, modelling and, not least, the geographical context (for more details see Debrezion et al., 2007 and Mohammad et al., 2013). In a meta-study drawing on 57 US studies, Debrezion et al. (2007) found that commuter railway stations show significantly higher effects on property prices than other types of rail such as heavy railway or metro stations. In a more recent meta-analysis, Mohammad et al. (2013) also found that the impact of commuter rail on land and property values is higher than that of light rail transit. That study also found that the impact of railways in Asian and European cities seems to be higher than in American cities. In previous research, both beneficial effects of transport as well as disutilities were included (Bowes and Ihlanfeldt, 2001; Kilpatrik et al., 2007). Previous studies have compared the impact of highways and rail systems on property prices (Seo et al., 2014).

What most studies fail to acknowledge is that the impact may differ across price segments. The effect of improved railway transportation on housing prices may not be equally important for all households. For the most expensive houses, rail transit may be of lesser importance than it is to middle income households. Some studies do mention this issue (Bowes and Ihlanfeldt, 2001; Debrezion et al., 2007; Nelson, 1999), but to our knowledge there is little research on this theme.

The purpose of this study is to explore the effects of commuter trains and highways on residential property prices, with a special focus on how effects vary across market segments. We specifically model the segmentation of sales prices for single-family houses and the influence of different explanatory variables on each segment using quantile

* Corresponding author.
*E-mail address:* helena.bohman@mah.se (H. Bohman).

[1] The authors gratefully acknowledge the financial support from K2 The National Knowledge Centre for Public Transport.

http://dx.doi.org/10.1016/j.jtrangeo.2016.09.003

regression techniques. In order to incorporate the spatial distribution of observations, we apply a two-stage spatial quantile regression (2SQR). There is also a methodological contribution from the study. In the model, two property-specific neighbourhood variables are developed through raster data in order to capture characteristics of the vicinity, thus generating measures that do not depend on administrative borders.

The geographic setting of this study is the Scania region in the south of Sweden and its regional commuter train system. The region is of interest for a couple of reasons. First, because the regional commuter train system has a strong position and is central for the functioning of the labour market. Commuting by regional train in the region almost tripled between 2000 and 2014 (Region Skåne, 2016). Second, most previous studies focus on smaller geographical areas and our aim is to provide an understanding for the regional context. Furthermore, no previous studies exist on this specific area. Fig. 1 presents an overview of the region, showing population density, the four largest cities, and railway and highway networks.

The paper is organized as follows. The following section provides a brief background to hedonic price models and the literature on the modelling of price segments through the use of quantile regression. This is followed by the results from a spatial lag model applied and adapted to quantile regression. The results confirm the expectation that preferences differ across different market segments.

## 2. Price segments and quantile regression

The changing consumption patterns across incomes are well established in demand studies. The Engel curve represents the fundamental microeconomic observation that the propensity to consume certain goods depends on the level of income of the consumer. Typically, the share of basic commodities decreases when income increases, whereas the share of luxury goods tends to increase with income. Roed Larsen (2006) found that the demand for transport differs across market segments. The share of income spent on petrol and public transport decreases as income increases, whereas air flights and leisure travel

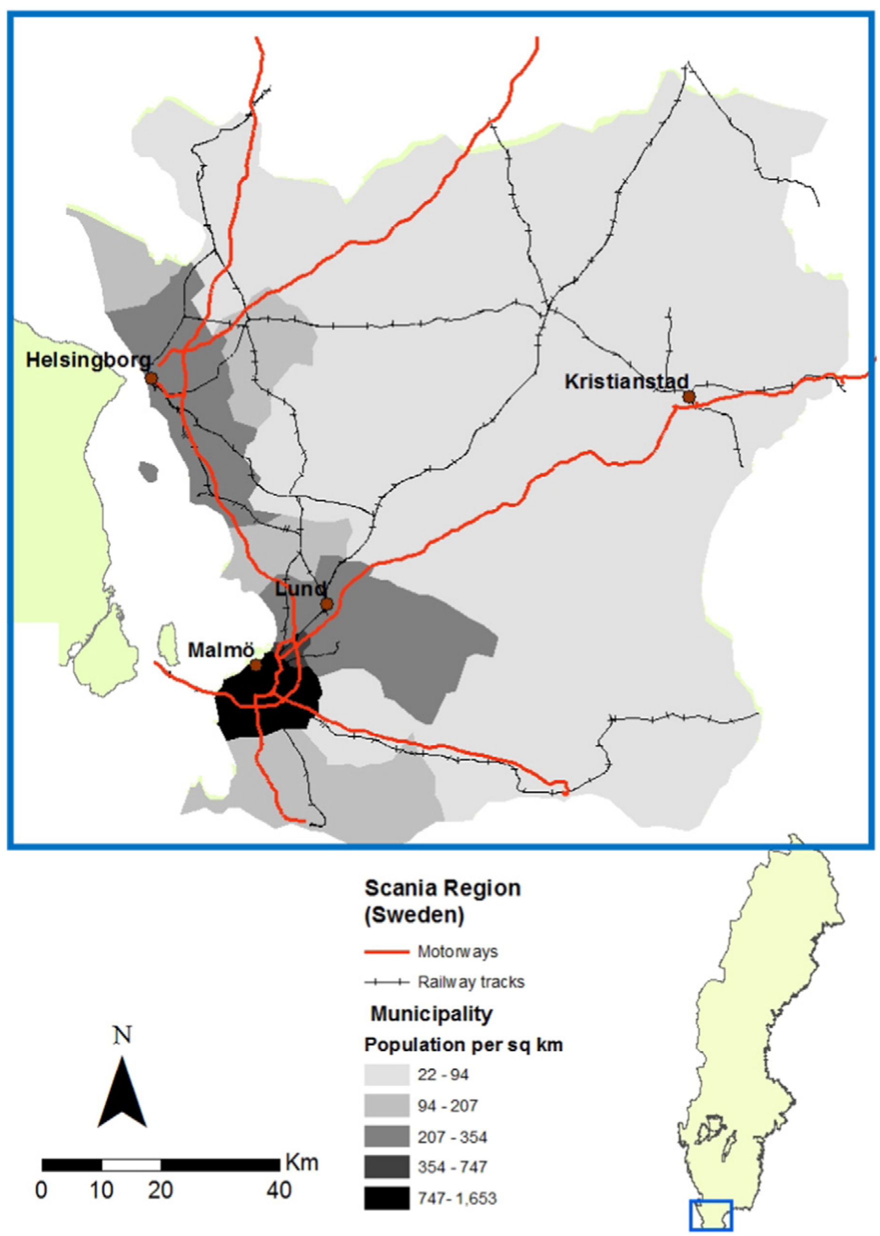


Fig. 1. Overview map of Scania.

can be considered luxury goods since their shares of household expenditure increase with income.

Given that housing is a highly heterogeneous good, it is useful to consider submarkets. A recent study by Baudry and Maslianskaia-Pautrel (2015) suggest that failure to consider heterogeneity may result in biased estimates, but there are different definitions of what constitutes a submarket. Bourassa et al. (2003) claim that the definition may vary depending on the focus of the study. Segmentation of the housing market is generally considered from a geographic (horizontal) perspective, whereas the focus of this study is on vertical segmentation, i.e. market segments that are due to consumer income. There are a few previous studies focusing on vertical segmentation, such as Farmer and Lipscomb (2010) and Liao and Wang (2012). Zietz et al. (2008) found that several attributes of a house, including square footage, number of bathrooms and age, vary across segments. In line with Baudry and Maslianskaia-Pautrel (2015), the aim of this study is to include vertical segmentation in order to find better estimates of housing values. The focus is on understanding how different price segments of the single-family house market may be explained differently by various factors.

The property market is a highly heterogeneous market, where different characteristics affect the values of a property. These characteristics range from attributes of buildings to neighbourhood characteristics but also regional characteristics. Hedonic price models express prices as a function of different attributes and has thus become a common tool to analyse real estate prices. The theoretical underpinnings of the hedonic price model were provided in Lancaster (1966) and developed by Rosen (1974). Hedonic price models can also be used to estimate non-market values such as green areas, noise, other pollutants, and proximity to waste sites.

In this study, we do however want to address an additional aspect of heterogeneity not frequently addressed in the literature, namely the variation in demand across income levels. One way to address the Engel curve for a market could be to divide the market into submarkets. As Heckman (1979) showed, however, truncating the dependent variable may cause biased estimates. Instead, quantile regression offers an alternative through which it is possible to use the full data set while at the same time allowing for nonlinearities across the dependent variable. Sections 2.1 and 2.2 will focus on quantile regression and its application in a geographical context.

### 2.1. Quantile regression and spatial data

Whereas ordinary least square (OLS) regression techniques estimate an average effect based on the mean value of the data, quantile regression is better suited to providing estimates that involve changes in the distribution of the explanatory variable. One of the merits of quantile regression is that the method provides more robust results with data exhibiting heteroscedasticity through allowing coefficients to vary with the spectrum of the dependent variable. Although similar to OLS, quantile regression minimizes the weighted sums of absolute residuals instead of minimizing the sums of squared residuals. Early work on quantile regression date back to the 1970s (e.g. Koenker and Bassett, 1978), and the method has since been applied in numerous fields (Koenker and Hallock, 2001), including consumer studies and demand analysis.

Deaton (1997) offers an introduction to quantile regression techniques applied to demand studies. He found that although the mean value differs little from the median value, there are systematic variations over quantiles supporting the Engel curve, with lower income groups spending a higher share of their income on food than higher income groups. Bayer et al. (2004) also found support for the hypothesis that households with different income levels value various aspects of housing differently. Similar to our study, Zietz et al. (2008) addressed price segmentation using quantile regression to estimate housing prices.

The application of quantile regression to spatial data in the previous literature is more limited (McMillen, 2013). One example is Liao and Wang (2012) who applied the method to Chinese housing data and found that the demand for quality of life, measured through access to parks, increased with income. They concluded that quantile estimates should be used together with mean estimates when studying hedonic prices. Another example is Kostov (2009) who applied quantile regression to agricultural land.

### 2.2. Model

A common method to control for spatial autocorrelation is the use of a special autoregressive (AR) model in which a weighted average of geographically close values of the dependent variable is added to the list of explanatory variables. The general AR model (McMillen, 2013) can be written as.

$$Y = \rho WY + X\beta + \varepsilon \tag{1}$$

where $Y$ is the dependent variable and $X$ a set of explanatory variables. $W$ is an n*n matrix specifying the spatial relationship between each value of $Y$ and its neighbours, and n is the number of observations. $\rho$ is a weight representing the strength of the spatial relationship between the dependent variable and nearby observations. We chose to estimate the weight matrix using normalized third-order queen contiguity estimation in the program GeoDa (Anselin et al., 2006).[2] The queen contiguity allows for shared borders as well as corner connections between spatially close observations. Third-order contiguity allows for a larger spatial area to be considered as having an impact on a specific observation. Since we study a region where spatial difference may be vast, we thought a higher order to be more appropriate.

However, applying a spatial AR model to quantile regressions is a more complex procedure than in the case of standard regressions, since instrumental variables are needed for $WY$ when estimating a regression for each quantile $\tau$ (for an overview see McMillen, 2013). The two-stage estimation of a spatial lag model used here is presented in Kim and Muller (2004) and Liao and Wang (2012). Liao and Wang (2012) refer to the model as a two-stage spatial quantile regression (2SQR). We estimate the following model for each quantile $\tau$:

$$P = \lambda_\tau WP + X\beta_\tau + \varepsilon_\tau \tag{2}$$

where $p$ is the log-transformed house prices, $X$ is a set of explanatory variables, and $\lambda_\tau$ and $\beta_\tau$ are to be estimated. Estimations are made in two steps. The first step is to estimate the spatially lagged $Wp$ against the spatially lagged exogenous variables $WX$ and $X$. The estimated value, $\widehat{Wp}$, is then substituted into the second-stage regression to control for spatial autocorrelation.

## 3. Data and model specification

The empirical setting for this study is the region of Scania in southern Sweden. The region covers 11,027 km$^2$ and has 1.3 million inhabitants. About 48% of the land is used for agriculture. There are formally three local labour markets. The neighbouring cities of Malmo (32,000 inhabitants) and Lund (117,000 inhabitants) are the centre of the largest labour market. The second largest labour market centres on the city of Helsingborg (138,000 inhabitants), while the city of Kristianstad is the centre for the third local labour market. Since the early 1980s a network of regional train lines has been developed to connect the smaller towns in the region with the larger cities. The commuter trains use the same tracks as the national trains. The network includes more than 60 commuter train stations, with smaller stations linking with a few relatively large stations.

There is great variation in the formulation of hedonic price models (see the meta studies of Debrezion et al., 2007 and Mohammad et al.,

[2] A software program developed by the GeoDa Center and its affiliates (R-Geo).

2013). To some extent, this can be explained by different national and local standards in data collection. There is also a large variation in estimation techniques when it comes to how distances are measured. It can also be argued that different housing markets may exhibit different characteristics due to cultural and environmental differences, and therefore variations in the models and resulting outcomes should be expected. Nonetheless, some general patterns can be identified in most previous studies. Typically, there are property-specific attributes such as size of the house and lot size, age of the house, the number of bedrooms or bathrooms. Secondly there are characteristics that refer more to the surrounding environment, such as the population size and local economy that may determine property values.

In this study, explanatory variables are divided into three different subgroups: property-specific attributes, neighbourhood-specific attributes, and transport-related variables. Details on all variables are presented in Table 1. The property-specific data includes all single-family house transactions in the Scania region during 2014 as collected through the National Land Survey (Lantmäteriet). A quality feature of the data is that all transactions are registered, rather than just data collected from realtors. The data quality is also a reason for the focus on housing rather than commercial real estate, since in the Swedish market commercial real estate is often sold packaged into companies and thus is not entered in the land register. The data set contains information on factors such as size, age, house type, the quality of housing, beach access and leasehold of land. Each house is separately geocoded at the street address in order to be able to calculate different distance variables, providing a precise level of property location.

**Table 1**

| Variable | Description | Source |
|---|---|---|
| **Dependent variable** | | |
| *PRICE* | Price (log transformation) of single-family houses, 2014 | NLS |
| **Property attributes** | | |
| *SIZE* | Floor area, defined as living area ($m^2$) | NLS |
| *LEASE* | Leasehold of land owned by municipality, 1 if house is leasehold 0 otherwise | NLS |
| *DETACHED* | Detached house, 1 if house is detached 0 otherwise | NLS |
| *AGE* | Age of building. Houses constructed before 1929 are marked as 1929. The stated age may be adjusted after major renovations or extensions | NLS |
| *ANTIQUE* | 1 if the year of construction was before 1940, 0 otherwise | |
| *BEACH* | Beach-side property, 1 if closer than 150 m to the waterfront, 0 otherwise | NLS |
| *QUALITY* | Quality index of the physical attributes of the building | NLS |
| **Transportation attributes** | | |
| *STATION_DIST* | Travelling distance using road network (km) | STA [a)] |
| *DEPARTURES* | Number of departures for the closest train station in 2014 | Skånetrafiken |
| *STATION_CLOSE* | Dummy variable where 1 indicates property within 200 m from station, 0 otherwise | STA [a]) |
| *HW_DIST* | Distance to nearest highway (hundred km) | STA [a]) |
| *HW_CLOSE* | Dummy variable where 1 indicates property within 200 m from highway, 0 otherwise | STA [a]) |
| *CBD_DIST* | Distance to nearest labour market centre (thousand km) | STA [a]) |
| *MALMO_LUND* | 1 if the closest station is in Lund or Malmo, 0 otherwise | STA [a]) |
| **Neighbourhood attributes** | | |
| *INCOME* | Average income, municipality level | SCB |
| *POP* | Population size, municipality level | Region Skåne/SCB |
| *INCOME_NEIGHBOUR* | Average income proxy based on raster data, vicinity of house | SCB |
| *POPDENS* | Population density calculated from raster data, vicinity of house | NLS |
| *COAST* | 1 if house located in coastal municipality, 0 otherwise | |
| *OSTERLEN* | 1 if house located in the Osterlen beach community, 0 otherwise | |

NLS = Lantmäteriet/National Land Survey (2015), SCB = Statistics Sweden (2015), STA = Swedish Transport Administration (2016).

[a] Calculated using GIS-based road network data.

The second subgroup, and the focus of this study, consists of transport-related variables. Property access to the nearest commuter train station is calculated as the road distance from each property to the closest open train station, using road network data (Fig. 2). We also account for access to highways, for previous research (e.g. Debrezion et al., 2007) indicates that introducing highway accessibility to the model may reduce the impact of train stations. Although commuting by train is an important means of transport, most commuting to work is still undertaken by car. Major roads are distinguished by including only E-roads.[3] Incorporating access to both highways and train stations will throw light on how the effect of commuter train access compares to the effect of highway access. Both variables are intended to reflect the positive effects of accessibility to train stations and highways.

Living in close proximity to either a station or a highway might incur negative externalities due to such things as noise pollution. For this reason, the model includes two dichotomous variables that reflect closeness to a station and closeness to a highway, using 200 m as a threshold. Bowes and Ihlanfeldt (2001) used a threshold of a quarter of a mile and found that properties within this range sell for less than properties situated beyond it. Li and Saphores (2012) found that the negative effect on sales prices is larger for the impact area of 100–200 m than for further distances.

To further account for some characteristics of the train stations, the number of departures from each station is included as an additional variable. Estimates were made using both passenger statistics and the number of departures in order to measure the traffic intensity of the stations. However, since these two variable formulations proved to be highly correlated, only the number of departures is included in the estimates below, since this should be less prone to measurement error.

The third group of variables is neighbourhood and municipal level characteristics. Two different geographic levels are used to capture socio-economic effects. At the more aggregate level, we use municipal level data. However, we also have access to finer raster level data, allowing for defining the surrounding geographical area of each property rather than depending on administrative boundaries. Income and population data are presented in raster format with a grid of 250 m by 250 m in more populated areas and 1000 m by 1000 m in more sparsely populated areas. The close neighbourhood is then defined within the radius of a kilometre from the property, providing unique measures for each property. This means that the close neighbourhood in more populated areas may contain up to four square kilometres (8 times 8 groups of 250 by 250 m), while in the more sparsely populated areas close neighbourhoods are defined by up to nine square kilometres (3 times 3 groups of 1000 by 1000 m). These neighbourhood level variables are then used to calculate nearby population density as well as income level. The close neighbourhoods are smaller than the administrative borders of the municipality and can overlap one another.

Both the municipal level and the neighbourhood level are included in the models since the geographic coverage may to some extent capture different types of information. For income, both geographic levels capture the mean income[4] for each specified geographic unit, although

[3] The international E-road network is a classification for roads in Europe developed by the United Nations Economic Commission for Europe (UNECE).

[4] For the raster data, only median income is available within each raster. The defined variable is therefore aggregated from this data and is thus a proxy of mean income within the specified neighbourhood.

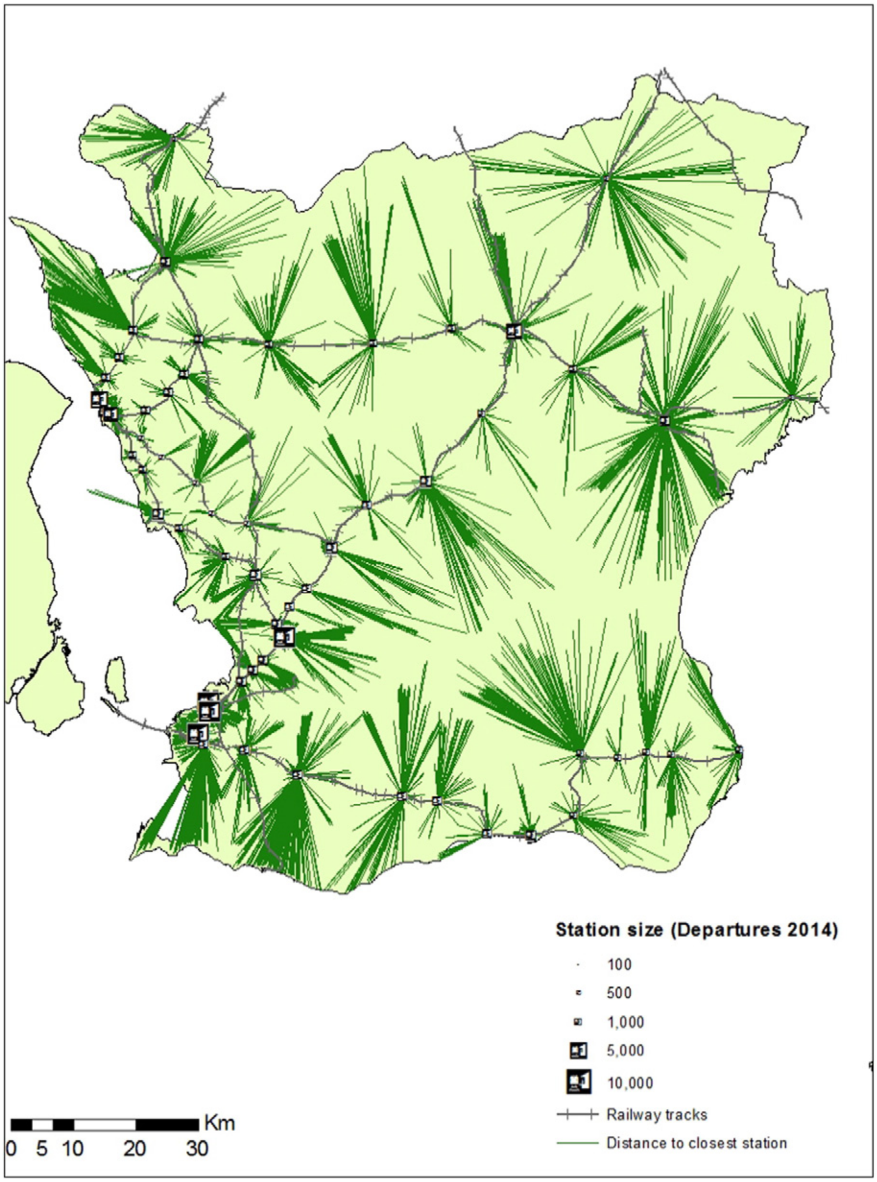


**Fig. 2.** Map of estimated proximity to the closest commuter train station for single-family home transactions. Note: Proximity is calculated through road network (illustration of distance 'as the crow flies' solely for visual purposes). Station symbols capture number of departures from each station in 2014.

the neighbourhood level allows for larger local variation. Bowes and Ihlanfeldt (2001) warn that including the median income in the model may indirectly result in an underestimation of the proximity effect, since railways may contribute to higher income levels due to better accessibility. However, we will still include these variables, while acknowledging that the real effect may be higher than the estimated effect. For population data, the two levels are defined so as to capture two different effects. The municipal level measures overall population, and so captures a market-size effect. At the local level, however, population density is thought to provide a measure of individual space. Although urbanization shows that people tend to value closeness to larger urban communities, larger individual space may still be an attractive feature.

In order to account for proximity to the centre of the functional labour market, two different measures are used. The administrative council of Scania describes the region as a multiple city (polycentric) region with three local labour markets. As a result of this, distances to the centre of largest city within each labour market are estimated. In addition to this, a dichotomous variable is introduced to capture whether the adjacent cities Malmo and Lund in the largest local labour market might have additional effects on the value of properties in the region.

The data set contains property transactions. The average property was sold for 2.3 million SEK and had a size of 130 square metres. The majority of the houses are detached. Two per cent of the houses are land lease. Land lease implies that the municipality owns the land but the resident, who pays for a long-term lease of the land, owns the building. The period of the lease is normally 26 to 100 years. Land lease housing has been a political instrument to increase home ownership at a relatively low cost, and transactions are registered the same way as for other single-family houses. More than two-thirds of the houses in the sample are detached houses. The age variable is calculated from the recorded year of construction. Although age is expected to affect prices negatively, houses built before the 1930s are often in the higher-price segments of housing. This may partly be explained by the low supply of housing from that era and by the fact that the houses still remaining are often high-quality buildings, whereas poorer housing is likely to have been replaced with newer buildings. There is also an aesthetic aspect in that many buyers find early twentieth century

**Table 2**
Descriptive statistics for dependent and explanatory variables.

| Variable name | Mean | Min | Max | Std. dev. |
|---|---|---|---|---|
| Price | 7.5 | 0 | 11 | 0.86 |
| Size | 130 | 17 | 2904 | 68 |
| Lease | 0.017 | 0 | 1 | 0.13 |
| Detached | 0.79 | 0 | 1 | 0.41 |
| Age | 51 | 1 | 85 | 24 |
| Antique | 0.23 | 0 | 1 | 0.42 |
| Beach | 0.026 | 0 | 1 | 0.16 |
| Quality | 30 | 8 | 56 | 4.5 |
| Station_dist | 7.6 | 0 | 38 | 7 |
| Departures | 1.9 | 0 | 5.8 | 1.5 |
| Station_close | 0.019 | 0 | 1 | 0.14 |
| Hw_dist | 0.04 | 0.00011 | 0.11 | 0.027 |
| Hw_close | 0.00059 | 0 | 1 | 0.024 |
| CBD_dist | 0.023 | 0.000025 | 0.08 | 0.017 |
| Malmo_Lund | 0.16 | 0 | 1 | 0.36 |
| Income | 265 | 232 | 372 | 31 |
| Pop | 8 | 0.72 | 32 | 9.5 |
| Income_neighbour | 0.49 | 0 | 5.6 | 0.58 |
| Popdens | 0.55 | 0 | 9 | 0.95 |
| Coast | 0.58 | 0 | 1 | 0.49 |
| Osterlen | 0.055 | 0 | 1 | 0.23 |

architecture attractive. In order to capture this non-linearity, a dummy variable is included for houses from this period.

As regards closeness to either a commuter train station or a highway, only 2% of properties are within 200 m of a train station, and the number of properties near highways constitutes far less than 1% in this data set. The average distance by road to a train station is slightly more than 7 km. Table 2 summarizes the different variables used in the empirical application of our model.

## 4. Results

The main purpose of this study is to provide new insights on how access to transport may affect income segments differently, seen from the perspective of property prices. Table 3 shows the findings from equation 2, applied to the 10th, 25th, 50th, 75th and 90th percentiles. Transport-related variables have been framed in the table.

Starting with the property-specific characteristics, size becomes slightly more important over the quantiles. The land lease variable is negative for all but one quantile, the 25th percentile, for which it turns out to be insignificant. This is in line with expectations. Detached houses yield overall higher prices, and the impact is consistently stronger for the higher-price segments. In order to account for the age of the building, the hypothesis is a non-linear finding in the sense that older houses in general yield lower prices up to a certain age. However, for the older houses captured by the Antique variable and representing houses constructed before 1940, there is a price premium. This premium also increases for the higher-price segments, which is in line with expectations, since houses constructed before the 1930s often yield fairly high prices and this could be interpreted as a sort of amenity value. Proximity to the beach has a positive impact on all market segments, as does the QUALITY index. The quality variable impact is, however, reduced over the market segments, possibly since it is an aggregate measure and may not capture all aspects of quality.

The main focus of the study is, however, the impact of transport-related variables. Distance to a train station has a negative impact on price, but this effect is most important for the 10th percentile. This is in line with the Engel curve and previous research (Roed Larsen, 2006) stating that lower income group depend more on public transport. Fig. 3a plots the estimated coefficients across market segments for distance to the nearest train station, measured by the Station_dist variable. The plots are based on calculations for each decile in order to provide a more detailed pattern than the one presented in Table 3. A similar but much stronger pattern is revealed for distance to highways

**Table 3**
Results from two-stage spatial lag quantile estimates. Dependent variable: ln(Price). Unstandardized coefficients. Bootstrapped standard errors in parentheses. *significant at 10%, **significant at 5%; ***significant at 1%.

| | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 |
|---|---|---|---|---|---|
| Size | 0.000 | 0.001*** | 0.002*** | 0.003*** | 0.003*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Lease | −0.178*** | −0.059 | −0.121*** | −0.148*** | −0.146*** |
| | (0.065) | (0.038) | (0.030) | (0.043) | (0.041) |
| Detached | 0.037 | 0.050*** | 0.068*** | 0.101*** | 0.146*** |
| | (0.027) | (0.016) | (0.014) | (0.011) | (0.022) |
| Age | −0.006*** | −0.006*** | −0.005*** | −0.005*** | −0.004*** |
| | (0.001) | (0.001) | (0.000) | (0.000) | (0.001) |
| Antique | −0.027 | 0.086** | 0.138*** | 0.205*** | 0.210*** |
| | (0.043) | (0.037) | (0.023) | (0.018) | (0.021) |
| Beach | 0.566*** | 0.495*** | 0.643*** | 0.599*** | 0.536*** |
| | (0.094) | (0.056) | (0.042) | (0.032) | (0.051) |
| Quality | 0.025*** | 0.022*** | 0.017*** | 0.009*** | 0.004** |
| | (0.002) | (0.002) | (0.002) | (0.001) | (0.002) |
| Station_dist | −0.011*** | −0.004*** | −0.002* | −0.002** | −0.004*** |
| | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) |
| Departures | 0.034** | 0.031*** | 0.021*** | 0.036*** | 0.024* |
| | (0.016) | (0.011) | (0.007) | (0.006) | (0.013) |
| Station_close | −0.675*** | −0.365*** | −0.157*** | −0.135*** | −0.160*** |
| | (0.115) | (0.116) | (0.058) | (0.036) | (0.061) |
| Hw_dist | −3.641*** | −3.093*** | −1.483*** | −0.664** | −0.669** |
| | (0.824) | (0.384) | (0.264) | (0.314) | (0.321) |
| Hw_close | −1.276* | −0.462 | −0.078 | −0.154 | −0.203 |
| | (0.702) | (0.740) | (0.618) | (0.133) | (0.196) |
| CBD_dist | −2.873** | −3.243*** | −2.988*** | −2.129*** | −0.782 |
| | (1.349) | (0.982) | (0.680) | (0.770) | (1.220) |
| Malmo_Lund | 0.230*** | 0.158*** | 0.126*** | 0.106*** | 0.161*** |
| | (0.036) | (0.026) | (0.027) | (0.022) | (0.041) |
| Income | 0.006*** | 0.004*** | 0.004*** | 0.003*** | 0.004*** |
| | (0.001) | (0.000) | (0.000) | (0.000) | (0.001) |
| Pop | 0.013*** | 0.008*** | 0.008*** | 0.005*** | 0.003** |
| | (0.003) | (0.001) | (0.001) | (0.001) | (0.002) |
| Income_neighbour | 0.635*** | 0.465*** | 0.338*** | 0.184*** | 0.042 |
| | (0.064) | (0.045) | (0.021) | (0.036) | (0.040) |
| Popdens | −0.339*** | −0.230*** | −0.156*** | −0.062*** | 0.025 |
| | (0.036) | (0.023) | (0.014) | (0.023) | (0.029) |
| Coast | 0.245*** | 0.227*** | 0.192*** | 0.186*** | 0.184*** |
| | (0.061) | (0.012) | (0.016) | (0.017) | (0.018) |
| Osterlen | 0.512*** | 0.417*** | 0.356*** | 0.288*** | 0.229*** |
| | (0.081) | (0.054) | (0.039) | (0.041) | (0.069) |
| Spatial_lag | 0.157*** | 0.258*** | 0.293*** | 0.346*** | 0.368*** |
| | (0.042) | (0.025) | (0.024) | (0.034) | (0.046) |
| Constant | 3.332*** | 3.370*** | 3.476*** | 3.520*** | 3.709*** |
| | (0.253) | (0.153) | (0.159) | (0.184) | (0.253) |
| Observations | 8489 | 8489 | 8489 | 8489 | 8489 |
| Pseudo R2 | 0.31 | 0.34 | 0.32 | 0.29 | 0.27 |

(Fig. 3b). Although not totally comparable, the impact in this case also decreases (in absolute terms) across the segments. The pattern can be interpreted as indicating that not only public transport but also transport in general may matter little for the housing decisions of the richer households. For the properties closest to the station, there is a disamenity similar to that found in previous studies, but there is no similar effect for the highway.

When it comes to variables explaining the impact of the surroundings, the results indicate that both the income level at the municipal level and the more local level are positive and significant. The overall income level in the closer neighbourhood reveals a lower impact for the 90th percentile. A possible explanation for this is that the highest priced properties include several in areas that previously were not high income areas, such as fishing villages along the coast that have witnessed dramatic property price increases. All of the included houses are, however, permanent residences; no houses registered as summer homes are included in the study.

The results for the impact of the population size and density capture two different effects. There seems to be a bonus from being close to larger towns or cities, but there is also a negative effect from the population density of the close vicinity. Both effects decrease in absolute terms over

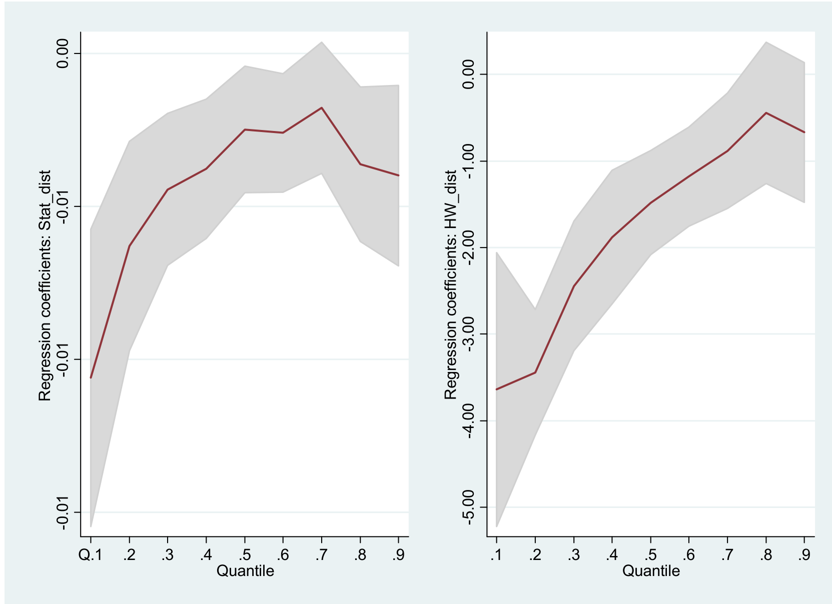


**Fig. 3.** a and b. Unstandardized coefficient plots from quantile estimates, with confidence intervals (see Azevedo, 2004).

the quantiles. The results can be interpreted as indicating a premium for proximity to large markets as well as a premium for individual space. These premiums remain even when controlling for lot size. Population size is measured at municipal level, whereas density is measured at neighbourhood level. At neighbourhood level, both income and population density show a stronger impact for the lower price segments. An interesting observation is that the two variables measuring characteristics of the nearest vicinity are insignificant for the highest market segment. The two local amenity dummy variables, Coast and Osterlen, reveal positive signs and decline over the market segments.

The Departures variable proxies the size of the node. However, the economic activity is concentrated at the centre of the local labour markets. CBD_dist displays a significant negative impact on house prices, with the exception of the highest segment. For all but the most expensive price segment, closeness to the centre of local labour market is essential for explaining property prices. The largest urban areas are the neighbouring cities of Malmo and Lund; whether they have some additional bearing on property prices may be captured by the dummy variable Malmo_Lund. This variable turns out to be positive and significant for all price segments. The effects seem to decrease with higher prices but to turn upwards again in the highest segment. We interpret this as an agglomeration effect: being close to an urban area matters, and it matters more if the urban area is relatively large.

Quantile regressions do not lend for $r^2$ calculations, and therefor pseudo $r^2$ estimates are presented for each model specification. The level of the pseudo $r^2$ are fairly low, something that we attribute to the large and diverse geographical area covered in the study. The pseudo $r^2$, although not really comparable to $r^2$, from these models seem to be generally lower than standard $r^2$ reported from OLS (see e.g. Zietz et al., 2008). The main purpose here is however to see if there is a difference between market segments, not aim to make predictions on property prices, and from that perspective low $r^2$ values are less of a problem.

The spatial lag coefficient increases over the market segments, implying that the unexplained spatial dependence coefficient increases with the higher price segments of the market. This differs from the finding of Liao and Wang (2012) who report a U-shaped pattern. There could be different explanations for this, geographical (richer households live in areas with geographical features that are not captured by this model) as well as socio-economic ones. Given that different data sets may capture different income spans, this could also influence the findings. Further research is however needed before any conclusions can be made. It is however important to note that adding the spatial lag variable does not in any major way change the impact of the other variables, compared to non-spatial specifications of the models that were performed but are not reported in this study.

## 5. Concluding discussion

The aim of this study has been to introduce the issue of market segments into the literature on the effects of regional commuter train traffic on single-family home property prices. The application of quantile regression in a hedonic price model provides a tool for better understanding of both housing and transport market segments. Understanding market segments may also help to understand who benefits from different investments and the development of markets. In this case, the focus is on single-family housing, and there may be reason to suspect that the results may differ when applied to other types of real estate, and also to other types of railway transportation. We thus expect to see similar studies in different contexts.

The findings suggest that the effect of access to commuter trains on residential house prices differs across market segments, with lower income groups benefiting more than richer groups. Given that the data used represen owner-occupied single-family housing, it seems reasonable to assume that this study does not include the poorest groups in the economy. These groups are in general excluded from home ownership and to a large extent rent their housing. The lowest quantile in this data set should therefore more properly be described as representing lower middle class groups, who are likely to also have access to a car. This is especially true for households in the countryside.

Public transport is often characterized as less important for high-income households (Roed Larsen, 2006), and the findings here are in line with this. The findings suggest that also accessibility to highways seems to be relatively more important to the lower price segments in the data set. We believe this can be interpreted along the same lines. For richer households, even if they tend to use the car relatively more, expenditures on transportation in total are, if not negligible, at least not a main concern in the choice of housing.

Proximity to CBD matters for all segments, except for the highest one. What is interesting is that there is a strong positive premium if the closest station is located in the largest urban area, Malmo-Lund. Here the effect is also significant for the 90th percentile. For a multi-centre region such as Scania, this is highly relevant, not least in understanding how we should define local labour markets and their relative importance given the proximity to other centres. It suggests that the main urban area has a stronger effect on all market segments than the other regional centres. In order to better understand the relative importance of urban areas, more research is however needed.

Two new measures of the neighbourhood were also introduced in order to capture effects at both neighbourhood level and at a more aggregate level. Municipal data are readily available but limit the analysis to administrative borders that do not always correspond well with socio-economic patterns. Raster data allow for measurements that are more flexible, for borders can be adapted to suit the focus of the study. In this study, we included one measure of the income level of the closest neighbours and one capturing population density. A first general result is that both geographical levels matter for the analysis, although none of the neighbourhood level variables are significant for the 90th percentile.

Summarizing our findings, we find that:

- Both distance to train stations and distance to highways have a significant negative impact on property prices, indicating that accessibility to commuter trains and highways matters for all market segments.
- Preferences differ across market segments. The effect of distance to a train station is highest for the 10th percentile, a segment that also reveals the strongest disutility from being too close to a train station.
- Proximity to the largest urban area, the Malmo-Lund region, is more important than proximity to the other centres. For multi-centre regions, of which there are many in Europe, this has a bearing on the relative importance of the different centres.
- Population density at a municipal level shows a positive impact on property prices, whereas the reverse is true at the neighbourhood level. Using variables not depending on administrative borders provides methodological flexibilities that strengthen the analysis.

## References


Andersson, D.E., Shyr, O.F., Fu, J., 2010. Does high-speed rail accessibility influence residential property prices? Hedonic estimates from southern Taiwan. J. Transp. Geogr. 18 (1), 166–174. http://dx.doi.org/10.1016/j.jtrangeo.2008.10.012.

Anselin, L., Syabri, I., Kho, Y., 2006. GeoDa: an introduction to spatial data analysis. Geogr. Anal. 38 (1), 5–22.

Azevedo, J., 2004. GRQREG: Stata module to graph the coefficients of a quantile regression. Statistical Software Components S437001. Boston College Department of Economics (revised 17 Mar 2011).

Baudry, M., Maslianskaia-Pautrel, M., 2015. Revisiting the hedonic price method in the presence of market segmentation. Environ. Econ. Policy Stud. 09. http://dx.doi.org/10.1007/s10018-015-0122-5.

Bayer, P., McMillan, R., Rueben, K., 2004. An equilibrium model of sorting in an urban housing market. NBER Working Paper Series 10865. http://dx.doi.org/10.3386/w10865.

Bourassa, S., Hoesli, M., Peng, V., 2003. Do housing submarkets really matter? J. Hous. Econ. 12 (1), 12–28. http://dx.doi.org/10.1016/S1051-1377(03)00003-2.

Bowes, D.R., Ihlanfeldt, K.R., 2001. Identifying the impacts of rail transit stations on residential property values. J. Urban Econ. 50 (1), 1–25. http://dx.doi.org/10.1006/juec.2001.2214.

Chen, C.-L., Hall, P., 2011. The impacts of high-speed trains on British economic geography: a study of the UK's InterCity 125/225 and its effects. J. Transp. Geogr. 19 (4), 689–704. http://dx.doi.org/10.1016/j.jtrangeo.2010.08.010.

Deaton, A., 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. Johns Hopkins, Baltimore (ISBN: 0-8018-5254-4).

Debrezion, G., Pels, E., Rietveld, P., 2007. The impact of railway stations on residential and commercial property value: a meta-analysis. J. Real Estate Financ. Econ. 35 (02). http://dx.doi.org/10.1007/s11146-007-9032-z.

Debrezion, G., Pels, E., Rietveld, P., 2011. The impact of rail transport on real estate prices: an empirical analysis of the Dutch housing market. Urban Stud. 48 (5), 997–1015. http://dx.doi.org/10.1177/0042098010371395.

Farmer, M., Lipscomb, C., 2010. Using quantile regression in hedonic analysis to reveal submarket competition. J. Real Estate Res. 32 (4), 435–460.

Geurs, K.T., van Wee, B., 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. J. Transp. Geogr. 12 (2), 127–140. http://dx.doi.org/10.1016/j.jtrangeo.2003.10.005.

Heckman, J., 1979. Sample selection bias as a specification error. Econometrica 47 (1) (pre-1986, Jan 1979).

Kilpatrik, J., Throupe, R., Carruthers, J., Krause, A., 2007. The impact of transit corridors on residential property values. J. Real Estate Res. 29 (3), 303–320.

Kim, T.-H., Muller, C., 2004. Two-stage quantile regression when the first stage is based on quantile regression. Econ. J. 7 (1), 218–231.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica Jan 46 (1), 33–50.

Koenker, R., Hallock, K.F., 2001. Quantile regression. J. Econ. Perspect. 15 (4), 143–156. http://dx.doi.org/10.1257/jep.15.4.143.

Kostov, P., 2009. A spatial quantile regression hedonic model of agricultural land prices. Spat. Econ. Anal. 4 (1), 53–81. http://dx.doi.org/10.1080/17421770802625957.

Lancaster, K., 1966. A new approach to consumer theory. J. Polit. Econ. 74 (2), 132–157. http://dx.doi.org/10.1086/259131.

Li, W., Saphores, J.D., 2012. Assessing the impacts of freeway truck traffic on residential property values: a Southern California case study. Transp. Res. Rec. 2288, 48–56. http://dx.doi.org/10.3141/2288-06.

Liao, W.-C., Wang, X., 2012. Hedonic house prices and spatial quantile regression. J. Hous. Econ. 21, 16–27. http://dx.doi.org/10.1016/j.jhe.2011.11.001.

McMillen, D., 2013. Quantile regression for spatial data. Springer Briefs in Regional Science http://dx.doi.org/10.1007/978-3-642-31815-3.

Mohammad, S., Graham, D., Melo, P., Anderson, R., 2013. A meta-analysis of the impact of rail projects on land and property values. Transp. Res. A Policy Pract. 50, 158–170. http://dx.doi.org/10.1016/j.tra.2013.01.013.

National Land Survey, 2015. Single-family House Transactions for 2014.

Nelson, A.C., 1999. Transit stations and commercial property values: a case study with policy and land-use implications. J. Public Transp. 2 (3), 77–95.

Pagliara, F., Papa, E., 2011. Urban rail systems investments: an analysis of the impacts on property values and residents' location. J. Transp. Geogr. 19 (2), 200–211. http://dx.doi.org/10.1016/j.jtrangeo.2010.02.006.

Region Skåne, 2016. Resandeutveckling på tågen. ([http://utveckling.skane.se/siteassets/samhallsplanering/dokument/tagresande-skane-per-station-2000-2014.pdf], accessed June 1, 2016).

Rietveld, P., Bruinsma, F., 1998. Is Transport Infrastructure Effective? Transport Infrastructure and Accessibility: Impacts on the Space Economy. Springer, Berlin (ISBN 978-3-642-72232-5).

Roed Larsen, E., 2006. Distributional effects of environmental taxes on transportation: evidence from Engel curves in the United States. J. Consum. Policy 29, 301–318. http://dx.doi.org/10.1007/s10603-006-9009-y.

Rosen, H., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. J. Polit. Econ. 82 (1).

Seo, K., Golub, A., Kuby, M., 2014. Combined impacts of highways and light rail transit on residential property values: a spatial hedonic price model for Phoenix, Arizona. J. Transp. Geogr. 41, 53–62. http://dx.doi.org/10.1016/j.jtrangeo.2014.08.003.

So, K.S., Orazem, P.F., Otto, D.M., 2001. The effects of housing prices, wages, and commuting time on joint residential and job location choices. Am. J. Agric. Econ. 83 (4), 1036–1048. http://dx.doi.org/10.1111/0002-9092.00228.

Statistics Sweden, 2015. Socio-economic Data at Regional Level.

Swedish Transport Administration, 2016. National Road and Railway Database.

UN-Habitat, 2016. Urbanization and Development: Emerging Futures (ISBN 978-92-1-133395-4).

Zietz, J., Zietz, E.N., Sirmans, G., 2008. Determinants of house prices: a quantile regression approach. J. Real Estate Financ. Econ. 37, 317–333. http://dx.doi.org/10.1007/s11146-007-9053-7.