DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper

Households' willingness to pay for access to outdoor recreation

An application of the house price method using spatial quantile regressions

Cathrine Ulla Jensen

2016 / 09

IFRO Working Paper 2016 / 09

Households' willingness to pay for access to outdoor recreation: An application of the house price method using spatial quantile regressions

Author: Cathrine Ulla Jensen

JEL Classification: R20, Q26, Q51

Published: October 2016

See the full series IFRO Working Paper here:
www.ifro.ku.dk/english/publications/foi_series/working_papers/

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 25
DK 1958 Frederiksberg DENMARK
www.ifro.ku.dk/english/

HOUSEHOLDS' WILLINGNESS TO PAY FOR ACCESS TO OUTDOOR RECREATION

AN APPLICATION OF THE HOUSE PRICE METHOD USING SPATIAL QUANTILE REGRESSIONS

Author:

Cathrine Ulla Jensen

Affiliation:

Department of Food and Resource Economics, Faculty of Science, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg Copenhagen, Denmark

Contact details
Email: cuj@ifro.ku.dk
Phone +45 3533 3677

Affiliation
Department of Food and Resource Economics
Faculty of Science, University of Copenhagen
Rolighedsvej 23
1958 Frederiksberg Copenhagen, Denmark

Acknowledgements:

This paper benefited from funding through the project "Benefits from investing in urban life and urban qualities", funded in a collaboration among the municipalities Aarhus, Albertslund, Brøndby, Frederiksberg, Gladsaxe, Glostrup, Herlev, Hvidovre, Ishøj, Københavns, Lyngby-Taarbæk, Rødovre, Skanderborg and Vallensbæk and the Capital Region of Denmark, the Ministry of Environment and Food, developers Kildebjerg Ry A/S and By & Havn I/S and the University of Copenhagen.

Abstract

This paper investigates how household demand for access to nature varies across a Danish housing market. I use conditional quantile regressions to estimate the implicit price for a change in nature area conditional on the home price. If there are systematic differences in the willingness to pay for nature across the house price distribution, a conditional mean will systematically under- or over-predict the impact for certain priced homes. In addition, this heterogeneity can be of interest to policy-makers as a potential indicator of the distributional profile associated with a given policy. This study investigates this question by employing both standard spatial econometrics and spatial quantile regressions. I find that nature in the vicinity of the home is perceived as an amenity across the entire market. This result is robust to different spatial controls. What is not robust is the size and the profile of the price premium. There is a large correlation between the income level within a neighborhood and the level of outdoor recreation. Controlling for unobserved quality through fixed effects reveals that the price premium increases with prices, but when controlling for the general price level using the trade price of neighboring homes (a lag), the price premium becomes constant. Controlling for local neighborhood affluence and unobserved quality on a larger scale yields the same results as the spatial lag term but with a more robust model due to the absence of endogeneity. The paper offers a discussion of this discrepancy and relates it to the more general discussion of controlling for spatial dependence.

Keywords: Hedonic pricing, Spatial econometrics, Quantile regression, Pooled cross-section, Green space

INTRODUCTION

Public nature areas provide a range of services to people living in their vicinity. Living close to a nature area provides easy access to jogging, observing wildlife and other outdoor recreational activities within walking distance of the home. This good is not directly marketed, and as the development of land intensifies, it can easily be forgotten in favor of more tangible goods. The value of access to outdoor recreation or different forms of green space has been the subject of a number of hedonic valuation studies; see, e.g., Tyrväinen & Miettinen (2000), Morancho (2003), or Lake, Lovett, Bateman, & Day (2000) or, for more recent applications, Sander, Polasky, & Haight (2010) or Panduro & Veie (2013).

A central limitation among these studies is that they assume that the conditional mean adequately describes the price drivers across the whole price distribution. A new branch within the hedonic literature is exploring the dimensions of this heterogeneity in preferences across the market by assessing how attributes may be valued differently at different points on the conditional price distribution. Coulson & McMillen (2007) investigate how different physical attributes affect the value of a home conditional on the price of the home relative to all other homes in the market. They find significant and substantial variation, in parallel to Zietz, Zietz, & Sirmans (2008) and Levitt & Syverson (2008).

A way to understand this variation is to frame it as the existence of submarkets. A market consists of a number of overlapping submarkets that are all connected through the market mechanism. Very few households would consider both a home in the lowest percentile and one in the highest percentile of the price distribution when choosing their home. Even so, each household considers a different and overlapping segment of the entire housing supply within the overall market, and thus, the submarkets are still united under one larger market (Farmer & Lipscomb, 2010; Lipscomb & Farmer, 2005). In a policy context, the relevant question is as follows: Does the implicit price vary across submarkets to an extent that it matters for the decision maker?

Environmental (dis)amenities also affect expensive homes differently than they affect average or cheap ones. Kuethe & Keeney (2012) investigate how agriculture and waste management facilities affect nearby housing, and they find that negative externalities matter more for higher priced homes and that some negative externalities matter only for homes priced over the median. They find that conclusions drawn solely on the conditional mean models would either overestimate the effect for lower priced homes and underestimate the effect for higher priced homes or falsely conclude no effect. Chasco & Le Gallo (2015) find similar patterns when they investigate the price premium for noise and air pollution in the city of Madrid.

In Changsha, China, 40,000 homes were built and traded for the first time between 2008 and 2009. Liao and Wang (2012) study how different types of green space are valued across the price distribution when these homes were traded for the first time. They find considerable and significant variations in implicit prices for a number of attributes across the conditional price distribution. For natural parks, they find that a 100-m increase in distance to the nearest natural park results in a price change of -0.4 % to + 0.3% for the 10^{th} to the 90^{th} percentile with a conditional mean estimate of 0.07%. For urban parks, they find a price premium for a 100-meter increase of -0.1% to -0.4% with a conditional mean of -0.2%. By not censoring, they assume that the effect is constant regardless of the spatial scale, e.g., a move from 100 to 200 meters results in the same percentage change as a move from 2000 to 2100 meters. If households sort in relation to greenspace, then on average, lower priced homes will be situated further away from green space than higher priced homes.

A major concern in the application the hedonic method is the spatial nature of the data. If one aspect of unobserved quality captured by the error term is correlated with either households' choice of nature or just the level of nature in an area, then the estimate of nature may be biased. In other words, there exists an omitted variable bias. The value of a house consists of so many aspects that in practice, it is impossible to control for all of them. Even if a study included all relevant quality measures, the researcher would have to adequately relate the price and the quality measure in order to avoid a misspecification (von Graevenitz & Panduro, 2015). Omitted variables, misspecification and systematic mismeasurement are all potential sources of spatial autocorrelation or dependence. Both can lead to inefficient or even biased results (Anselin, 2010). It is complex problem, and there may be multiple omitted spatial processes at a number of different spatial scales (von Graevenitz & Panduro, 2015), where both the scale and importance are unknown to the researcher.

Liao & Wang (2012) control unobserved quality across 5 districts through spatial fixed effects and by including weighted neighborhood prices known as a lag term, in parallel to Zietz et al. (2008). Others include only neighborhood prices, e.g. Kuethe & Keeney (2012), and others control for variation by including neighborhood demographics. Weighted neighborhood prices describe a process where a price change in the first home spills over to all its neighbors, which again spills over to all their neighbors, including the first home. A change will ripple across the whole housing market back and forth until the spillover is so small that it is negligible. The key question is whether these spillover effects have any welfare implications. Small & Steimetz (2012) argued that researchers should think about what the lag term captures and use this to guide how spillover terms are evaluated. A spillover would imply that in addition to the utility from their own increase in a given attribute level, they enjoy utility from an increase in an attribute level from one of their neighbors. Another interpetation, which I argue, that the lag term captures that neighboring house prices are used as a proxy for neighborhood quality by buyers and sellers (Small & Steimetz, 2012). For a good introduction to spatial econometrics with weight matrices, please see Anselin (2002).

A more technical problem is that the lag term introduces endogeneity because prices act as both dependent and explanatory variables. With regards to quantile regressions, the literature offers two solutions. Kim & Muller (2004) solve it by instrumentation in a procedure similar to 2SLS. In the first stage, instruments for weighted prices are constructed using spatial lags of the exogenous explanatory variables. These instruments are subsequently used in the second part of the estimation, where the dependent variable is regressed on the exogenous explanatory variables and the predicted value of the weighted prices from the first stage. It is the approach used by Liao and Wang (2012), Zietz et al., (2008) and Kuethe & Keeney (2012)

In addition to the fixed effect and lagged neighborhood prices, I include the income level of the neighbors. I assess the potential of avoiding weighting, endogeneity and ambiguous interpretation and capture the unobserved quality by the income level of the neighbors.

In the next section, I describe the theoretical background. Then, I describe the study area, data sources and variables, followed by the empirical framework. Then, I present the results and discuss the caveats, followed by a brief conclusion.

THEORETICAL FRAMEWORK

The hedonic pricing method, or the house pricing method, uses the market for housing to identify household demand for goods bought as a part of housing. The method is attributed to Rosen (1974), who describes how to identify household-specific demand functions for housing goods. Hedonic pricing in relation to housing has been used to value numerous environmental goods consumed along with housing. In the following, I discuss aspects essential for this paper. For a more thorough and general description of the theoretical foundation or a discussion of the founding literature within environmental economics, please see Palmquist (2005).

Under the assumption of weak separability in housing characteristics and all other goods, the demand for each characteristic is independent of the prices of all other goods. The market for housing is assumed to be in equilibrium, where each household occupies its utility maximization housing bundle, and prices are market clearing. Hedonic price function h maps the relationship between market price p_i of home *j* and qualities X_j and e_j:

$$P_j = h(\mathrm{X}_j, e_\mathrm{j}) \; (1)$$

A household can consume either housing or a Hicksian composite good, c. X is a vector of observed housing characteristics, which captures the aspect of quality observable both for the researcher and the household. Unobserved quality e is a scalar that captures all aspects of housing quality; it is observable only to agents within the housing market and not to the researcher. Housing quality includes the quality of the house itself, e.g., the number of rooms or age, or neighborhood characteristics describing various qualities of the surroundings, e.g., the tax level, the quality of the nearest school, shopping opportunities or environmental goods

Households are rational utility maximizers that choose their preferred housing bundle, j*, given their income y and preferences. Each household thus faces the following maximization problem:

$$\max_{\mathrm{x,q,c}} f\left(\mathrm{X_j}, e_\mathrm{j}, \mathrm{c}\right) \mathrm{s.t.}\, \mathrm{y} = \mathrm{h}(\mathrm{X_j}, \mathrm{e_j}) + \mathrm{c} \; (2)$$

For a housing bundle j* to be the utility maximizing choice for a given household, the following first-order condition must hold for each continuous good $\mathrm{x}_{\mathrm{j}*,\mathrm{k}}$:

$$\frac{\delta \mathrm{f}(\mathrm{X}_{\mathrm{j}*}, e_{\mathrm{j}*}, \mathrm{c})/\delta \mathrm{x}_{\mathrm{jk}}}{\delta \mathrm{f}(\mathrm{X}_\mathrm{j}, e_\mathrm{j} \mathrm{c})/\delta \mathrm{c}} = \frac{\delta \mathrm{h}(\mathrm{X}_{\mathrm{j}*}, \mathrm{e}_{\mathrm{j}*})}{\delta \mathrm{x}_{\mathrm{j,k}}} \; (3)$$

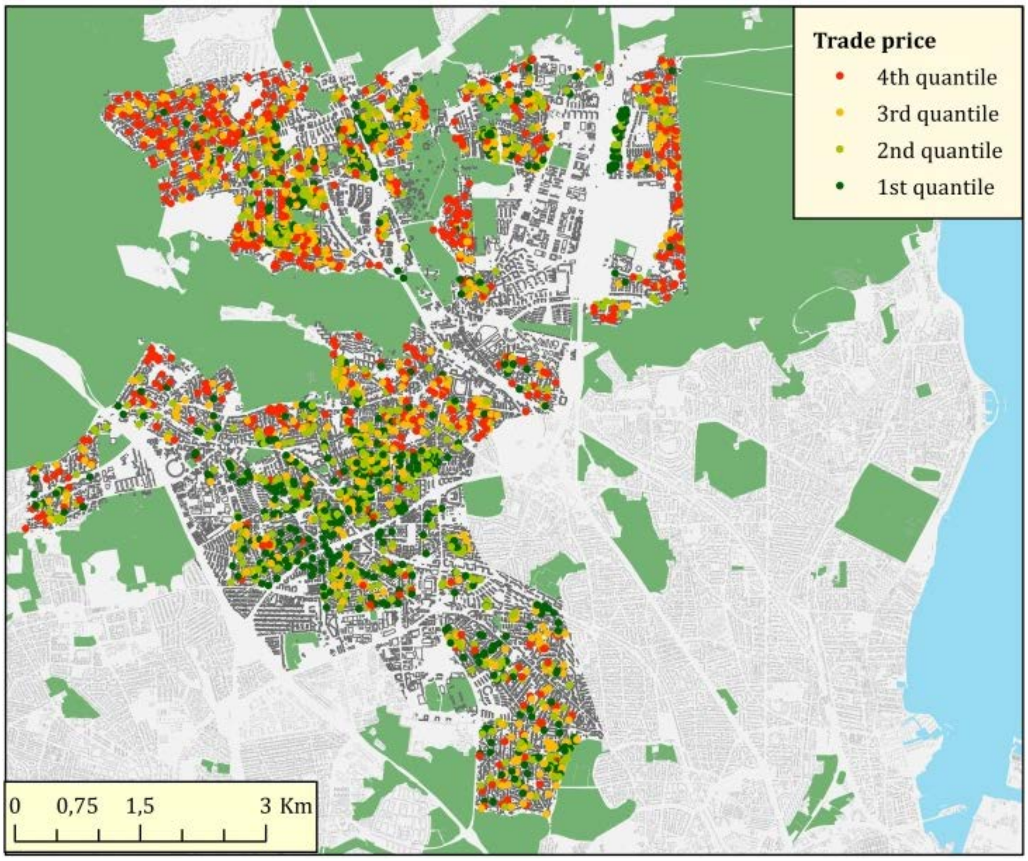
Households balance attributes to the point where the marginal rate of substitution equals the marginal costs. Given the assumption on market equilibrium, we just need to identity the right side of the first-order condition, namely, the implicit price of the good.

Formally, the hedonic pricing method consists of two stages. In the 1st stage, the researcher estimates the pricing function and through that an implicit price for each attribute. The 2nd stage recovers the demand functions. The present study concerns only the 1st stage and the importance of allowing the pricing function to capture heterogeneity.

DATA

The analysis is based on 2,445 single detached homes situated in the suburbs just north of Copenhagen, Denmark; see figure 1

Figure 1: The survey area and prices



The homes were traded between 2007 and 2010. Compared to the suburbs west and south of Copenhagen, the northern suburbs are characterized by homes that are priced higher on average and have higher amenity levels, measured based on green space. On average, the inhabitants are well-educated, are wealthier and make more money compared to those in the neighboring suburbs. Within the area itself, there is also sorting.

The highest priced homes are more abundant in the north and along the borders of the nature areas, depicted both by figure 1 and the descriptive statistics across the price distribution in table 1.

Households sort themselves across space in relation to affluence and the quality of housing. The house pricing function includes a number of control variables that measure aspects of housing quality, with regards to both the house itself and the surrounding neighborhood. Table 1 provides descriptive statistics for the variables included in the reported pricing function for the full sample and each 20th percentile. From the table alone, it is evident that when households choose their home, they bundle housing quality and amenity value. Higher priced homes are on average larger and situated in areas with a high level of nature and little road nuisance. In addition, these homes are situated in areas with high-income neighbors. Not controlling adequately for this correlation between green space and neighborhood quality could result in attributing WTP for other neighborhood qualities to WTP for access to outdoor recreation because the level of these two attributes is highly correlated.

Table 1: Mean Across the Price Distribution

	0-20%	20-40%	40-60%	60-80%	80-100%	Full sample
Price (1000 EUR)	268.66	371.63	441.29	522.21	729.63	466.70
Area (sqm)	102.42	106.88	117.21	133.61	162.39	124.50
Bathrooms	1.11	1.16	1.21	1.27	1.58	1.26
Rooms	4.07	4.25	4.54	4.92	5.55	4.66
Garden (sqm)	508.20	535.28	617.59	720.08	942.04	664.59
Tile roof	0.45	0.48	0.58	0.54	0.52	0.51
Built after 1999	0.02	0.02	0.01	0.01	0.03	0.02
Built 1875-1910	0.07	0.06	0.06	0.07	0.11	0.07
Rebuilt in the 1980s	0.02	0.03	0.05	0.05	0.07	0.05
Rebuilt in the 2000s	0.03	0.02	0.04	0.04	0.10	0.05
Road proximity max(0, 200- distance)	47.76	29.20	22.45	21.83	15.99	27.45
Nature within 800 m (ha)	18.02	21.71	26.89	36.35	49.22	30.44
Median income within 400 m (1000 EUR)	439,15	462.80	485.42	537.82	610.28	507.13
Postal 2830	0.19	0.22	0.31	0.39	0.39	0.30
Postal 2860	0.24	0.21	0.21	0.13	0.08	0.17
Postal 2880	0.24	0.20	0.13	0.12	0.12	0.16
Traded in 2008	0.19	0.24	0.24	0.21	0.24	0.22
Traded in 2009	0.38	0.28	0.21	0.23	0.12	0.24
Traded in 2010	0.32	0.31	0.25	0.28	0.30	0.29

SPATIAL VARIABLES

Spatial variables are incorporated into the housing price function by two different measures: censored proximity and density. The final specification includes only proximity to large roads, nature density and neighborhood income, but I test for a number of covariates using both measures.

Qualities that do not vary within the postal code will be captured by the fixed effect, e.g., distance to the capital or distance to the beach.

The measure for outdoor recreation is hectares of forest, meadow and lake within 800 meters. Parks are different in the context of recreation because they are more groomed and thus offer a different service than nature areas (Panduro & Veie, 2013). I include parks as a separate quality in the initial pricing function, but I find no significant effect. This is likely due to the number of homes in the sample situated near a park. The same is the case for beaches. Forest, meadows and lakes are biotopes with public access in Denmark . This makes these areas available for everyday outdoor recreation in a peri-urban area in a Danish context. The cut-off distance of 800 meters is defined empirically. I test a number of distances (e.g., 200 m, 400 m), walking distances and shortest distances. I find a nature area within 800 meters to outperform all other measures. The reason that beeline distance outperforms walking distance could be that not all small paths and trails that people use are included in the GIS network. Then, the Euclidian distance becomes a better proxy for the true perceived distance. A cut-off distance of 800 meters is decided based on empirics and seems reasonable. Danish studies based on surveys find the use of a nature area to decrease rapidly beyond a distance of 1000-m walking distance from home to site (Schipperijn et al., 2010; Toftager et al., 2011). The median income is calculated as the median income of all households within a given radius less the households in the home in question. This means that the measure is exogenous of the house price. I test distances from 200 m to 1200 m in 200 m intervals. I choose 400 m and a linear specification based on model performance.

DATA SOURCES

Housing data are extracted from the Danish Registry of Buildings and Housing, which contains structural information on all dwellings in Denmark. In addition, the registry contains information on the exact location of each property, making it possible to calculate a range of spatial variables describing relevant spatial public goods, such as distance to railways or roads, which can supply both easy transport and noise pollution for each home using R (R Core Team, 2015) and ArcGIS 10.2.1 (ESRI 2011, 2015). The spatial data are supplied by the Danish Geodata Agency based on the Kort10 database (The Danish Geodata Agency, 2011).

The demographic data describe households living the area in 2011 and include a number of variables. Due to the sensitive nature of individual-level demographic data, they are spatially blurred using a raster mosaic of 100*100 meters and then subsequently refined and matched to individual properties by Geomatic A/S. Another caveat is that the analysis covers trades in the period 2007 to 2010, but that period occurred before the period of the demographic data. I therefore assume that the general demographic composition within the area did not change radically between 2007 and 2011[1].

EMPIRICAL APPLICATION

The conditional mean regression describes relations at the mean of the sample, whereas the conditional quantile regression describes relations at any given percentile in the distribution.

I estimate a semilog pricing function:

$$\ln(p_i) = \beta \cdot X_i + \theta \cdot year + \delta \cdot postal + \lambda \cdot W \cdot ln(p_i) + \propto \cdot\, incNeighbors + \varepsilon_i \quad (4)$$

The price of home i is p_i. The log-linear specification is flexible and allows the marginal effect of each independent variable to vary with the level of the dependent variable (Cropper, Deck, & McConnell, 1988). This means that a constant coefficient across the price distribution corresponds to a price premium that increases with the house price in absolute terms. The existing literature offers little theoretical guidance on how to specify the functional form. The findings of Cropper et al. (1988) are often referred to in favor of the log-linear specification because they find it to be more robust to

[1] I use median income in 2011 within 400 meters as an explanatory variable in the pricing function and a house sample from 2007-2010. The median is relatively stable to changes compared to the mean. From 2007 to 2010, 2442 single detached homes were traded across the study area. In 2011, the population consisted of over 50.000 individuals. Given the ratio between trades and population, the neighborhood composition is unlikely to have changed dramatically.

misspecifications given there are no omitted variables. I report the results from a log-linear specification based on the above finding and model performance measured by explanatory power.

X is a vector of observed qualities including outdoor recreation. I test a range of spatial control variables using different spatial ranges and functional forms in the pricing function, as suggested by von Graevenitz & Panduro (2015). The choice between covariates and their spatial scale is based on an evaluation of estimates from a pricing unction estimated at the conditional mean. The cut-off is determined by considering model fit and parameter efficiency over a range of selected proximity cut-off values and density radiuses. The dataset covers 4 years, and the general price trend is controlled for by a vector of trade year dummies "*year*".

I control for omitted spatial variables in three ways- I impose a spatial fixed effect, $\delta_{l,}$ on postal code. The Danish postal code system was implemented in a time when post was distributed across the country by trains that originated from the capital, and thus, the delineation follows historical train lines. Postal codes still tell something about the neighborhood. For instance, insurance companies let the price premium of car and home insurance vary by postal code due to differences in risk. The general idea behind spatial fixed effects is that it nets out all the observed and unobserved factors that are in common within the area but different between areas, e.g., a local tax level or school quality. The appropriate spatial scale to capture a "neighborhood" is not known a priori. I have tried different meaningful fixed effects (school districts, postal codes and municipalities). In choosing the right neighborhood scale, I have to balance the desire to limit the risk of omitted spatial bias against the desire to identify the pricing function in the tails of the price distribution. As the spatial scale becomes finer, the number of supporting data points becomes scarcer. This limits the possibility for identification in the tails of the price distribution.

I include all neighbors in the median income, not only those residing in a home traded in the sample. The model presented in the results section includes median income within a 400-meter beeline (excluding households living in traded homes). I have tested the median income in 200-meter intervals up to 1000 meters. I have chosen 400 meters based on model performance. A distance of 400 meters is walkable, and one interpretation is the variable captures the social and economic capital of the immediate surroundings. Either the social and economic capital of household living further away does not matter for the household, or the desirability of those households is captured by the fixed effect on postal codes.

$\lambda W ln(p_i)$ is the lag. W is an NxN weighting matrix whose diagonal elements are zero and off-diagonal elements specify the neighborhood relationship between each observation in the sample. I use the 15 nearest neighbors in the matrix and inverse distance as weights. This specification has a nice intuition, where neighbors' impact decreases with distance but at a decreasing rate; a move from 5 to 10 meters decreases the impact more than a move from 300 to 305 meters. Second, the specification means that the number of neighbors included in the matrix drives the results very little because of the distance decay (Liao & Wang, 2012). I test 5 to 30 in neighbors in intervals of 5, and indeed, the results vary very little.

The lag term describes a process where a price change in the first home will spill over to all its neighbors, which again will spillover to all their neigbours, including the first house. This means that a change will ripple acros the whole housing market back and forth until the spillover is so small that it is neglible. LeSage (2008) offers some intepreations of the lag term, but von Graevenitz and Panduro (2015) note that not all are consistent with hedonic theory. A spillover could capture a process of neighborhood genitrifcation, where a change induces certain income groups to move in and other to move out and thereby changes the whole price level of the neighborhood. A large relocation of households within a housing market might induce the market equilibirum to change, which would imply a new pricing function. Another intepretation that does not violate the equilibrium assumption is that the spillover in reality is an information effect, where the price of recently sold neighboring homes are used by housebyuers and sellers as an indiciation of neighborhood quality. A problem is then that W includes all homes in the sample and not just past trades or listed homes around the trade date.

The spillover effect implies that a change in the level of an attribute for one house will capitalize in the entire housing market. LeSage (2008) structures this ripple effect by differentiating between the direct effect, defined as the the marginal effect of a one-unit change in the dependent variable for the one property, and the indirect effect, defined as the marginal effect of a one-unit change in the dependent variable for the neighbor(s). The sum of the two is then the total effect of a change.

Small & Steimetz (2012) argue that the researcher should think about what value the lag term captures, and this should guide the decision of which WTP to report. If the spillover is an information effect, then the measure should be the direct effect.

Another, more technical problem with the lag is that it introduces endogeneity; prices act as both a dependent and an explanatory variable, making OLS ill-suited for estimation.

Turning to the conditional quantile regressions,

$$ln(p_i) = \beta_\tau \cdot X_i + \theta_\tau \cdot year + \delta_\tau \cdot postal + \lambda_\tau W ln(\mathrm{p_i}) + \propto_\tau \cdot incNeighbors + \varepsilon_i \quad (5)$$

The difference between (5) and (4) is subscript τ in β_τ, λ_τ and $\propto_\tau$, which refers to the specific τ^{Th} percentile. The only difference between (4) and (5) is that the quantile regression describes any given relationship for conditional percentile τ instead of the conditional mean.

There is no analytical solution behind the quantile regression (QR) estimates. QR minimizes the weighted sum of absolute residuals, where the weight depends on the quantile and the individual observation on the sign of the residual. At the median, the weight is symmetrical, with half of the residuals positive in the solution. The percentile that is conditioned can be changed by changing the weights, where more asymmetrical weights will lead to the tails of the distribution. Formally, the weighting function for the τ^{Th} quantile is defined as $\rho_\tau(w) = w\big(\tau - I(w < 0)\big)$ for $\tau \in (0,1)$, where $I(\cdot)$ is an indicator function. The estimation is set up as a minimization problem and solved by iteration (Koenker and Hallock 2001) using the Quantreg package (Koenker, 2013). The quantile regressions that suffer from endogeneity are estimated using the approach developed by Kim & Muller (2004) using the McSpatial package (McMillen, 2013), and the results are reported with bootstrapped standard errors. All regressions are estimated using R (R Core Team, 2015). The conditional mean models without lagged prices are estimated using a standard OLS and reported with robust inference. The endogeneity problem in the mean lag model is solved in the spdep-package in an ML-estimation procedure; please see (Anselin et al., 2012) for an elaboration.

RESULTS

The section starts with the results for the conditional mean models and a discussion of their robustness to spatial autocorrelation. Then follow the results for outdoor recreation and a discussion of how much the demand for outdoor recreation varies across the price distribution.

I test a range of covariates including both other structural attributes that capture the individual quality of the home (e.g., number of rooms, year built, renovations before 2000, type of roof, basement) and neighborhood amenities (e.g., road and railway noise, access to parks, coast, distance to railway station). Distances are tested with both walking distance and beeline and different truncations. The choices are based on model fit using the conditional mean models. The final models include 17-19 covariates. Across all models, the coefficient signs are in line with expectations, e.g., the price increases the living area.

Table 2 shows a subset of the conditional mean models. At the conditional mean, just above 40% of the variation in price is explained.

Table 2: Pricing functions (a subset of the full regression results)

Model	**FE (1)**	**FE+income (2)**	**FE+lag (3)**	**FE+lag+income (4)**
Intercept	3.1051***	3.0936***	2,9956***	2,985***
	(0.1146)	(0.1141)	(0,1118)	(0,1116)
Log (living area)	0.5847***	0.5764***	0,5848***	0,5766***
	(0.0255)	(0.0252)	(0,0247)	(0,02480)
Postal code 2830	0.0227	0.0041	0,0221	0,0039
	(0.0141)	(0.0146)	(0,0145)	(0,0154)
Postal code 2860	-0.1010***	-0.0944***	-0,1008***	-0,0944***
	(0.0180)	(0.0182)	(0,0173)	(0,0173)
Postal code 2880	-0.1148***	-0.1185***	-0,1132***	-0,1168***
	(0.0193)	(0.0194)	(0,0176)	(0,0176)
Neighborhood income level (1000 EUR/year)		0.0012***		0.0012***
		(0.0003)		(0.0003
Lagged prices			0.0195***	0,0194***
			0.0033	(0,0033)
Outdoor recreation within 800 m (ha)	0.0019***	0.0016***	0,0018***	0,0016***
	(0.0002)	(0.0002)	(0,0002)	(0,0002)
Adjusted/ pseudo R^2	0.43	0.43	0.44	0.44

AIC	889	879	856	846

Note: Standard errors are HAC for models (1) and (2); *p<0.1; **p<0.05; ***p<0.01.

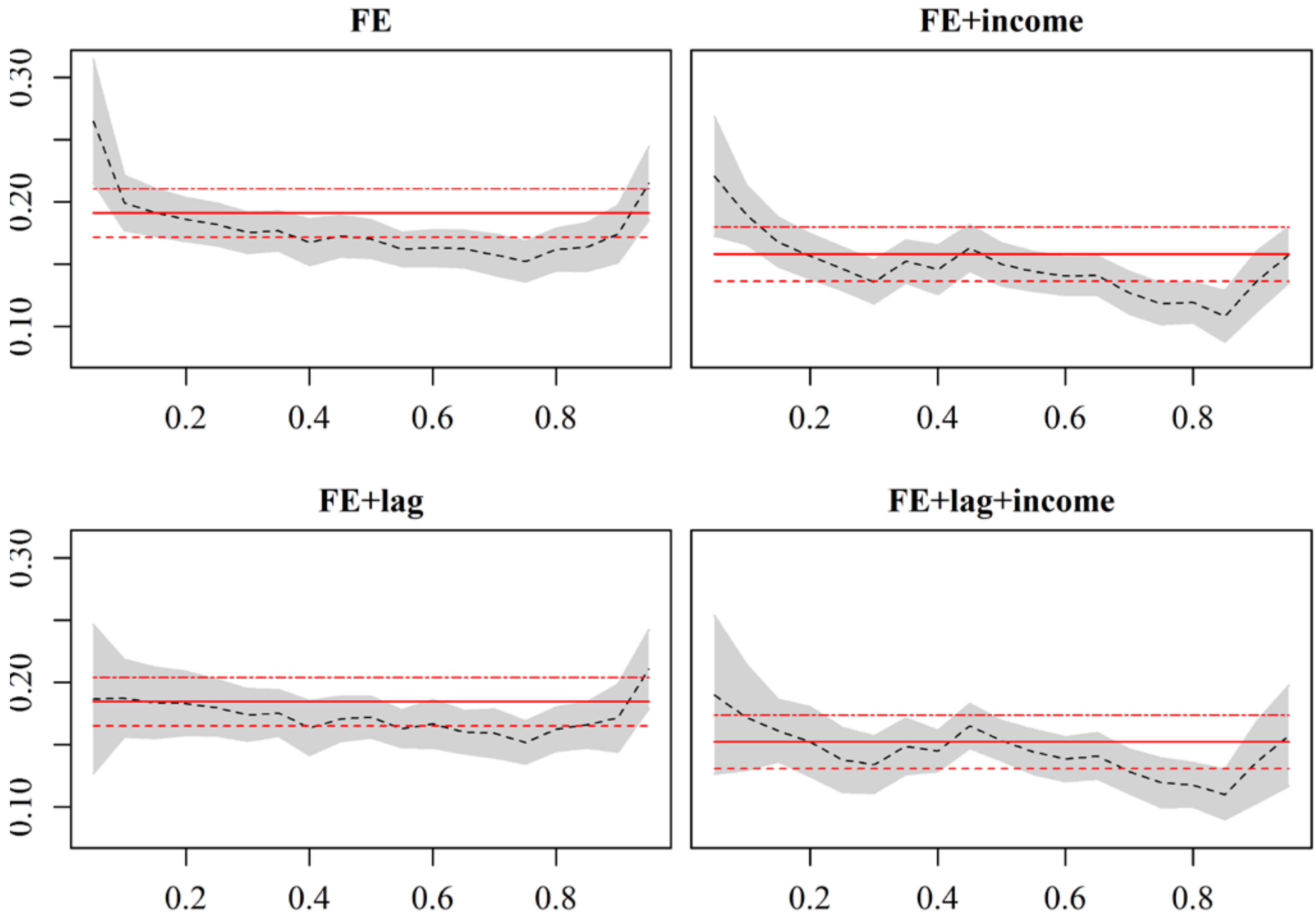
The first model includes only spatial fixed effects; the second, spatial fixed effects and median neighborhood income; the third, spatial fixed effects and distance-weighted neighborhood prices (lag); and the fourth, all three spatial controls. In the following interpretation, I assume that the lag term captures the informational value of neighboring homes and not the actual welfare that should be attributed to green space.

The price premium for outdoor recreation is between 0.16-0.19% (+-0.02%) for each additional hectare within 800 meters. Just based on the estimate, this corresponds to 74,672- 88.673 EUR per ha for the mean priced home (466,700 EUR). The average amount of outdoor recreation in the sample is 30 ha, so a 10% increase corresponds to -266,019- 224,016 EUR, a difference of more than 40.000 EUR. These examples are based on the coefficient estimates, and all have a standard error of 0.0002. It should be noted that only the two OLS models produce estimates that are significantly different from each other. The models with neighborhood income all cover 0.16%, and the FE+lag covers it at the lower bound. Including neighborhood income as a control decreases the estimated premium for outdoor recreation as well as the effect of weighted neighborhood prices. The neighborhood quality captured by income level and by price level is not overlapping; the two capture an omitted spatial process at two different scales. Neighborhood income takes out the unobserved quality captured by outdoor recreation, and this does not change with the inclusion of weighted prices. In conclusion, the implicit price for outdoor recreation at the mean is 0.16% (+-0.02%). The question is how well this represents other homes in the market.

Quantile regressions offer an insight into the heterogeneity across the price distribution, but it comes at the cost of simplicity. The quantile regressions are run on each 5-percentile, resulting in 19 estimated coefficients for each parameter for each model run. Therefore, the results are presented using figures.

The results for access to outdoor recreation are shown in figure 2; the remaining results are in the appendix. Each conditional mean result is presented in red by the estimated coefficient as a full line and standard errors by two dotted lines. The quantile regression is run on each 5th percentile, resulting in 19 estimated coefficients for each parameter in the pricing function. Each plot shows the price distribution at the x-axis and the implicit price premium in percentage of the home price at the y-axis. The estimated coefficient is connected by a black dotted line, and the standard errors are outlined by a grey polygon. Note that the same estimated coefficients in relative value are smaller in absolute value for the lower quantiles compared to the higher quantiles (i.e., 0.4% of 1 is less than 0.2% of 3).

Figure 2: Implicit price for outdoor recreation (in percent)



All four models predict a positive price premium across the entire price distribution; regardless of the price of the home, access to outdoor recreation holds amenity value. One more hectare within 800 meters results in a price premium between 0.3% and 0.8%. The model with the most control for unobserved quality reports a price premium of 0.1 to 0.2%. Controlling for unobserved quality outweighs the variation in price premiums across the distribution.

The two figures on the left include neighborhood income, and the two bottom figures include weighted prices. Regressions that include lagged prices need to be robust to endogeneity. In general, they produce larger standard errors. OLS is more efficient than ML, but in the presence of spatial autocorrelation, it can produce biased estimates and standard errors that are too small because observations are treated as independent. Errors are not independent if there is unobserved quality correlated with covariates through space. The larger standard errors in both tails are likely the result of fewer support points. Comparing the left and right sides, it is evident that controlling for neighborhood income reveals a different distribution profile.

There is a high correlation between the concentration of resourced households and access to nature areas. Without neighborhood income, the premium is relatively constant, with ambiguous or higher prices in the tails. In contrast, when neighborhood income is controlled for, there is a general decline across the distribution until the 90th percentile, followed by a sharp increase.

In the context of a CBA, the relevant measure is WTP in absolute terms. Table 3 shows the implicit price in absolute terms for the two models with all three spatial controls.

Table 3: WTP for one more ha of outdoor recreation (EUR)

Percentile	**House price**	**Price premium** (Quantile regression)	**Price premium** (Mean regression)	**Difference** (percent)
10%	289,204	498	440	-13%
20%	336,284	513	512	0%
30%	371,257	498	565	12%
40%	403,541	585	615	5%
50%	440,532	676	671	-1%
60%	477,523	662	727	9%
70%	520,567	670	793	16%
80%	574,050	677	874	23%
90%	672,568	917	1024	10%

Based on the model "FE+lag+income"

The difference between an estimate based on the conditional mean compared the relevant conditional percentile fluctuates between -13% and 23%. A conditional mean estimate overestimates the benefits of creating a new peri-urban nature area for homes priced above the median and underestimates it for cheap homes priced under the 20th percentile. In particular, the price premium for homes priced around the 3rd quantile is exaggerated by the conditional mean model; their willingness to pay is only ¾ of the WTP predicted by the conditional mean model. Both the conditional mean and the quantile regression model predict a WTP for outdoor recreation that increases with the price of the home. Surprisingly, the quantile regression shows that this difference is actually smaller than predicted by the conditional mean. Relative to the total price of the home, homebuyers shopping for cheap housing are willing to allocate a larger share of the total purchase to outdoor recreation. This might be a matter of supply and demand within each segment. Few small and low-quality homes are located places with a high level of nature in the vicinity, as shown in table 1. This means that the supply of cheap homes with a high level of green space are scarce; this drives the price of that combination upwards.

It is possible that the quantile regression actually recovers variation in nature quality and not variation in demand. One would expect that cheaper housing, on average, is situated close to low-quality green spaces compared to higher priced homes or at least not vice versa. If this were the case, then the price premium would increase across the distribution. This is only the case for the homes priced at the 90th percentile or above. These homes are, on average, situated closer to the forest, and it could be that this higher WTP is due to either better access or a spillover from a view of a lake or a forest. I also test "shortest distance", but it is outperformed by density measured by goodness of fit. Even so, it can still be that the upward kink at the 90th percentile in reality is not only a higher demand for outdoor recreation but also a demand for a nice view.

The standard errors in the left tail of the distribution could also indicate larger variation, but it could just as well be the lower number of support points that decrease precision. If households sort in relation to the quality of the nature area, then it would be not the variation in quality but the demand for access to outdoor recreation that generated the variation in prices. The highest priced homes are generally closer to a nature area, but when the scale is 800, the composition is relatively mixed. This means that it in general, these nature areas are the same ones to which lower priced and all other homes have access.

CONCLUSION

There are differences in the valuation of access to outdoor recreation across the price distribution. A conditional mean exaggerates the variation in WTP for outdoor recreation across the price distribution. In fact, households that shop for median-priced homes are willing to spend a larger part of their total housing budget on more hectares of outdoor recreation compared to households that shop for houses priced above the median. These differences call for caution when the recreational benefits for the neighborhood are valued if the neighborhood deviates from the distribution of homes on which the implicit prices are based. This study also revealed that households in the northern suburbs of Copenhagen that have high access to outdoor recreation generally also live in affluent neighborhoods. This correlation introduces a high risk of confusing the demand for outdoor recreation with the demand for resourced neighbors. I find that controlling for unobserved quality outweighs the variation recovered by the quantile regression. I could reduce the bias of the reported implicit prices more by focusing on unobserved quality compared to variation in implicit prices across the price distribution.

REFERENCES

Anselin, L. (2002). Under the hood Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, *27*, 247–267.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, *89*(1), 3–25.

Anselin, L., Assunção, R., Berke, O., Blanchet, G., Blankmeyer, E., Carvalho, M., … Rogerbivandnhhno, M. R. B. (2012). Package " spdep ." *R-Cran*, 1–204.

Chasco, C., & Le Gallo, J. (2015). Heterogeneity in Perceptions of Noise and Air Pollution: A Spatial Quantile Approach on the City of Madrid. *Spatial Economic Analysis*, *10*(3), 317–343.

Coulson, N. E., & McMillen, D. P. (2007). The dynamics of intraurban quantile house price indexes. *Urban Studies*, *44*(8), 1517–1537.

Cropper, M. L., Deck, L. B., & McConnell, K. E. (1988). On the Choice of Funtional Form for Hedonic Price Functions. *The Review of Economics and Statistics*, *70*(4), pp. 668–675. Retrieved from http://www.jstor.org/stable/1935831

ESRI 2011. (2015). ArcGIS Desktop. Redlands, CA: Environmental Systems Research Institute.

Farmer, M. C., & Lipscomb, C. A. (2010). Using Quantile Regression in Hedonic Analysis to Reveal Submarket Competition. *Journal of Real Estate Research*, *32*(4), 435–460. Retrieved from https://search-ebscohost-com.ezp.lib.unimelb.edu.au/login.aspx?direct=true&db=bth&AN=57226935&site=ehost-live

Kim, T.-H., & Muller, C. C. (2004). Two-stage quantile regression when the first stage is based on quantile regression. *Econometrics Journal*, *7*, 218–231. http://doi.org/10.1111/j.1368-423X.2004.00128.x

Koenker, R. (2013). Quantreg: quantile regression. *R Package Version*, *5*.

Kuethe, T. H., & Keeney, R. (2012). Environmental externalities and residential property values: externalized costs along the house price distribution. *Land Economics*, *88*(2), 241–250.

Lake, I. R., Lovett, A. A., Bateman, I. J., & Day, B. (2000). Using GIS and large-scale digital data to implement hedonic pricing studies. *International Journal of Geographical Information Science*, *14*(6), 521–541. http://doi.org/10.1080/136588100415729

LeSage, J. P. (2008). An introduction to spatial econometrics. *Revue D'économie Industrielle*, (123), 19–44.

Levitt, S. D., & Syverson, C. (2008). Market distortions when agents are better informed: The value of information in real estate transactions. *The Review of Economics and Statistics*, *90*(4), 599–611.

Liao, W.-C., & Wang, X. (2012). Hedonic house prices and spatial quantile regression. *Journal of Housing Economics*, *21*(1), 16–27.

Lipscomb, C. A., & Farmer, M. C. (2005). Household diversity and market segmentation within a single neighborhood. *Annals of Regional Science*, *39*(4), 791–810. http://doi.org/10.1007/s00168-005-0020-z

McMillen, D. P. (2013). *Quantile Regression for Spatial Data*. *Media*. Springer Briefs in Regional Science.

Morancho, A. B. (2003). A hedonic valuation of urban green areas. *Landscape and Urban Planning*, *66*(1), 35–41.

Palmquist, R. B. (2005). Property value models. *Handbook of Environmental Economics*, *2*, 763–819.

Panduro, T. E., & Veie, K. L. (2013). Classification and valuation of urban green spaces - A hedonic house price valuation. *Landscape and Urban Planning*, *120*, 119–128.

R Core Team. (2015). *R: A Language and Environment for Statistical Computing* (Vol. 3.2.0). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Rosen, S. (1974). Hedonic Prices and Implicit Markets - Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*(1), 34–55. http://doi.org/10.1086/260169

Sander, H., Polasky, S., & Haight, R. G. (2010). The value of urban tree cover: A hedonic property price model in Ramsey and Dakota Counties, Minnesota, USA. *Ecological Economics*, *69*(8), 1646–1656.

Schipperijn, J., Ekholm, O., Stigsdotter, U. K., Toftager, M., Bentsen, P., Kamper-J??rgensen, F., & Randrup, T. B. (2010). Factors influencing the use of green space: Results from a Danish national representative survey. *Landscape and Urban Planning*, *95*(3), 130–137. http://doi.org/10.1016/j.landurbplan.2009.12.010

Small, K. A., & Steimetz, S. S. C. (2012). SPATIAL HEDONICS AND THE WILLINGNESS TO PAY FOR RESIDENTIAL AMENITIES*. *Journal of Regional Science*, *52*(4), 635–647.

The Danish Geodata Agency. (2011). *kort10* (Vol. 2011). Retrieved from http://download.kortforsyningen.dk/

Toftager, M., Ekholm, O., Schipperijn, J., Stigsdotter, U., Bentsen, P., Grønbæk, M., … Kamper-Jørgensen, F. (2011). Distance to green space and physical activity: a danish national representative survey. *Journal of Physical Activity & Health*, *8*(6), 741–749.

Tyrväinen, L., & Miettinen, A. (2000). Property prices and urban forest amenities. *Journal of Environmental Economics and Management*, *39*(2), 205–223.

von Graevenitz, K., & Panduro, T. E. (2015). An Alternative to the Standard Spatial Econometric Approaches in Hedonic House Price Models. *Land Economics*, *91*(2), 386–409.

Zietz, J., Zietz, E. N., & Sirmans, G. S. (2008). Determinants of house prices: a quantile regression approach. *The Journal of Real Estate Finance and Economics*, *37*(4), 317–333.