Urban Economics

Lecture 6: Hedonic model

Spring 2024
Tuukka Saarimaa

Outline

This lecture introduces the hedonic model

The model can be used to analyze how housing prices reflect

- The demand for different housing characteristics or attributes and
- The demand for local public goods and neighborhood amenities

The model can also be used for prediction purposes

The lecture does not follow the textbook

Modeling housing demand

So far, we have considered a highly simplified housing market where houses only differed in terms of size and the commuting cost associated with their location

In reality, housing is a multidimensional product traded in bundles and households value the many features of a house/dwelling

- Floor area, number of bedrooms, size of the yard, condition, structural quality, accessibility and other locational characteristics

Implicit and explicit markets

The notion of **implicit markets** denotes the process of production, exchange, and consumption of commodities that are primarily (perhaps exclusively) traded in "bundles"

- There is an implicit market for floor space or neighborhood characteristics

The **explicit market**, with observed prices and transactions, is for the bundles themselves, i.e. houses/dwelling

Hedonic regression model

In a **hedonic regression**, the aim is to estimate the **relationship between prices and product attributes** in a differentiated product market

The **regression coefficients or slopes** are commonly referred to as **implicit (or hedonic) prices**, which can be interpreted as

- The effect on the market price of increasing a particular product attribute, while holding the other attributes fixed
- E.g. how much does the price of a house increase with an additional bedroom, when holding other features of the house fixed?
- Or how much does the price of a house increase when travel time to CBD is decreased by 15 minutes, when holding other features of the house fixed? (*ceteris paribus*)

Hedonic model – theory

Rosen (1974, JPE):

- Given utility-maximizing behavior, the consumer's marginal willingness to pay for a small change in a particular attribute can be inferred directly from an estimate of its implicit price

Moreover, these implicit prices can be used to recover marginal willingness to pay functions

- Demand functions for use in valuing larger changes in attributes

Rosen's model was not about housing, but differentiated goods in general

- Supply/producer side of the model is very different in housing compared to many other goods

Applications

Demand for differentiated products: cars, PC's, houses

Price indexes

- How has the price of a constant quality house evolved?

Predicting the value of a house or a lot for tax purposes

Labor markets and wages

- In addition to wages, jobs differ in several other respects

Valuing environmental quality

- Are houses more valuable in cleaner environments and by how much?

Valuing local public goods, such as schools, accessibility, safety

Empirical implementation in housing market context

To estimate a hedonic regression model, you need data on individual house prices and attributes

- In some cases, more aggregated data can/has to be used

There is nothing special econometrically or statistically about estimating hedonic models in general

- It's just a regression model with all the usual problems in terms of selection bias, omitted variables, statistical inference etc.
- The problems depend on the research question at hand

Example of a dataset

Variable	Mean	Std. Dev.	Minimum	Maximum
Dwelling characteristics:				
Price (€)	172 585	113 801	60 000	1 500 000
Floor area (m^2)	53.7	25.4	11	362
Age (in years)	54.7	22.5	2	136
Condition (broker estimate):				
Good	0.55	0.50	0	1
Satisfactory	0.39	0.49	0	1
Poor	0.06	0.23	0	1
Situated at own lot	0.76	0.43	0	1
Elevator	0.32	0.47	0	1
Floor level	3.04	1.65	1	9
Total number of floors in the building	4.92	1.69	2	9
Maintenance charge (€/m^2/month)	2.90	0.75	0	8
Road distances (km):				
CBD	5.80	4.49	0.32	19.2
Nearest train or subway stop	1.25	0.80	0.002	5.83
Sea	1.23	1.31	0.01	7.28

Note: The data consist of 7,472 dwelling transactions from Helsinki in 2006 and 2007. All observations are from multi-storey buildings.

Variable	OLS, no FE		OLS, FE		2SLS, FE	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Dwelling:						
log(floor area)	0.854**	0.013	0.848**	0.005	0.845**	0.005
log(age)	-0.028*	0.013	-0.056**	0.005	-0.050**	0.008
Good	0.159**	0.009	0.150**	0.006	0.150**	0.006
Satisfactory	0.067**	0.009	0.066**	0.007	0.066**	0.007
Own lot	0.081**	0.012	0.034**	0.005	0.033**	0.006
Elevator	0.000	0.008	-0.002	0.004	-0.001	0.004
Floor level	0.016**	0.002	0.016**	0.001	0.016**	0.001
Number of floors	-0.007*	0.003	-0.012**	0.001	-0.012**	0.002
Maintenance charge	-0.010*	0.005	-0.011**	0.003	-0.011**	0.003
Distance CBD	-0.026**	0.004	-0.019**	0.005	-0.013*	0.007
Nearest train or subway stop	0.013*	0.009	0.018**	0.006	0.020**	0.006
Distance to sea	-0.013*	0.008	-0.029**	0.008	-0.019*	0.013
Neighborhood:						
Homeownership rate	-0.057**	0.008	-0.024**	0.003	-0.071	0.048
log(median income)	0.052**	0.015	0.027**	0.007	0.062*	0.035
Share of college educated	0.080**	0.017	0.042**	0.006	0.063**	0.022
Unemployment rate	-0.010**	0.007	0.001	0.003	-0.004	0.007
Share of pension h'holds	0.007	0.006	0.010**	0.003	0.025	0.015
Share of h'holds with children	-0.052**	0.011	-0.025**	0.005	-0.038**	0.014
Service jobs per capita	0.015*	0.006	-0.001	0.002	-0.002	0.003
Number of buildings	0.022**	0.008	0.014**	0.003	0.012**	0.004
Population	0.011	0.010	-0.025**	0.004	-0.023**	0.004
Mean floor area of units	0.030**	0.013	0.008	0.006	-0.006	0.016

250 m x 250 m grids (bracket spanning Neighborhood section)

Notes: The table reports results from linear models where the dependent variable is the natural log of transaction price. All models include quarter of sale dummies. Standard errors are robust to heteroscedasticity. ** and * indicate 1 and 5 percent significance levels, respectively.

Valuing local public goods

Valuing local public goods

The hedonic model can be used to analyze how much people "value" or are willing to pay for local public goods

- Quality of elementary schools, crime, air or environmental quality, public transport etc.
- The consumption of these goods is tied to residential location
- But there is no explicit market for these goods

Logic:

- Homebuyers implicitly purchase the right to consume a bundle of local public goods when they buy a house in a particular location
- Thus, the hedonic price function can be used to infer buyers' valuation for local public goods

A fable of two families

Consider two identical families: the Smiths and the Johnsons

Both families have school-aged children and need a bigger house

They find two neighboring houses that are for sale

The real estate agent gives the families a tour of the houses and they are identical

- Identical yards, access to same amenities (grocery store, bus stop, playground, local library etc.), same neighbors etc.
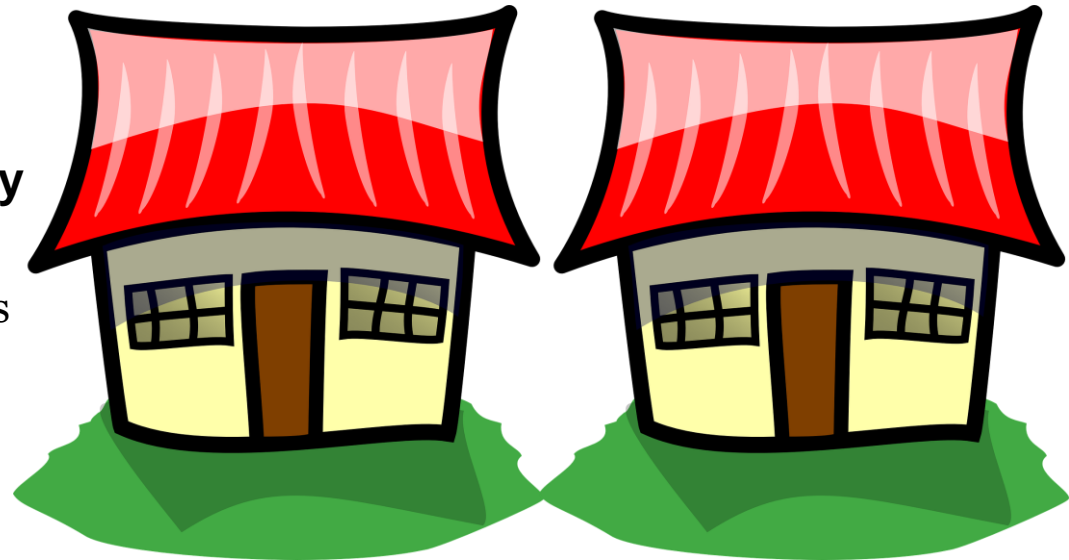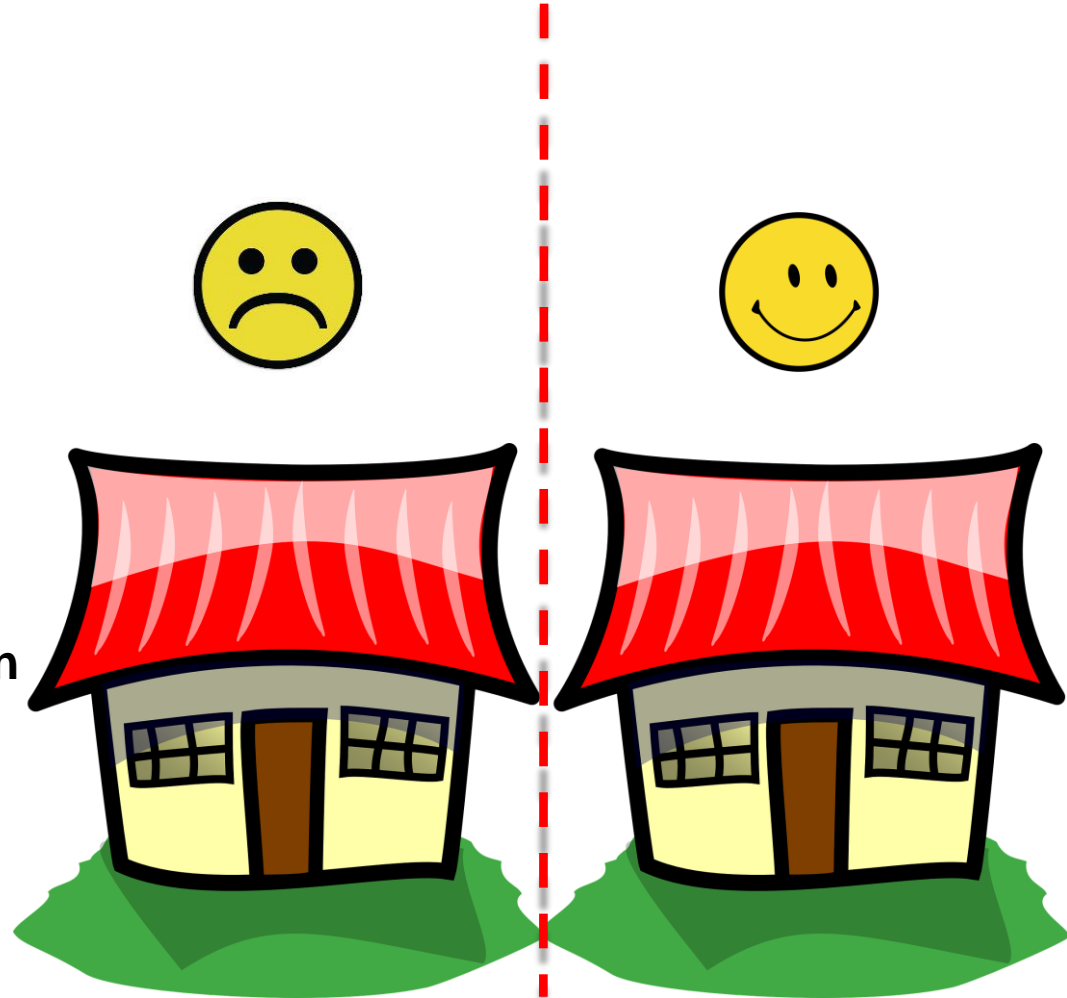
A fable of two families

However, the real estate agent reveals that the houses are in different school catchment areas

- From one, there is access to the best school in the municipality
- Both families now prefer this house

What do you think is going to happen when the families start bidding for the houses?

Bidding

After learning this, the families enter a bidding contest for the "better" house

The bids keep on increasing until the price difference between the houses is so high that the families are indifferent between the houses

- If the Smiths win the bidding contest, they get access to the better school (good for the Smiths)

- But pay more for the house and have less money for other stuff (vacations, restaurants, hobbies etc.) (bad for the Smiths)

- For Johnsons, this is the opposite

- Spatial equilibrium!
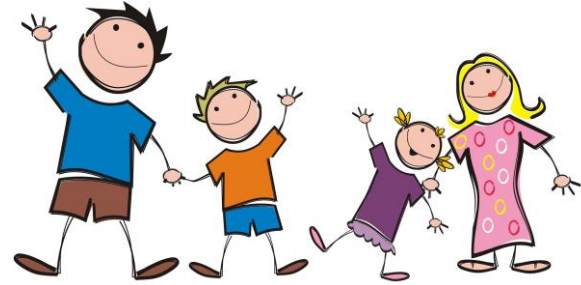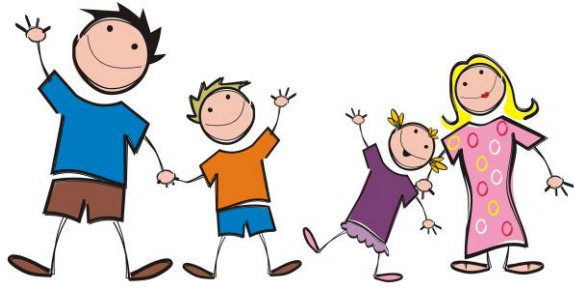
Are homebuyers in Helsinki willing to pay for access to good elementary schools?

Best Education Money Can Buy? Capitalization of School Quality in Finland

Oskari Harjunen*, Mika Kortelainen[†], and Tuukka Saarimaa[‡]

*Aalto University School of Business and City of Helsinki Executive Office – Urban Research and Statistics, Helsinki, Finland. e-mail: oskari.harjunen@aalto.fi, [†]VATT Institute for Economic Research, Helsinki, Finland and University of Manchester, Manchester, UK. e-mail: mika.kortelainen@vatt.fi and [‡]VATT Institute for Economic Research, Helsinki, Finland. e-mail: tuukka.saarimaa@vatt.fi

Abstract

By international comparison, Finnish pupil achievement is high and school achievement differences small. The Finnish education system is unusual also because there are no national testing programmes, and information on school quality measures is not publicly disclosed. Is school quality capitalized into house prices in this environment? Using a boundary discontinuity research design and data from Helsinki, we find that it is: 1 standard deviation increase in average test scores increases prices on average by 3%, which is comparable to findings from the UK and the USA. We argue that this surprisingly large effect is at least partly explained by the inelasticity of housing supply, as we use data from a densely populated urban area. We also show that the effect depends on local land supply conditions within the city and is highest in areas with inelastic supply. Furthermore, the price premium seems to be related to pupils' socioeconomic background rather than school effectiveness. (JEL codes: C21, H75, I20, R21)

School choice

Often school choice is based on residential location

- Sometimes this is an explicit rule: each housing unit is tied to a particular school through catchment areas
- Sometimes pupil attainment is freer, but residential location is still an important element in school choice (commuting costs etc.)

If school quality varies, we should expect this to be reflected in house prices

- Good schools can be accessed through the housing market
- Hypothesis: houses with access to better schools are more expensive (*ceteris paribus*)

Housing markets and segregation

1. **Neighborhoods within a city differ in terms of quality**

    - Accessibility, amenities, architecture, safety etc.

    - High-quality neighborhoods are a scarce commodity

2. **Households differ in terms of incomes and economic resources more generally**

3. **As their income increases, people consume nicer things**

    - This includes nicer residential neighborhoods, which can be purchased through the housing market

    - Housing prices serve as tickets to different neighborhoods
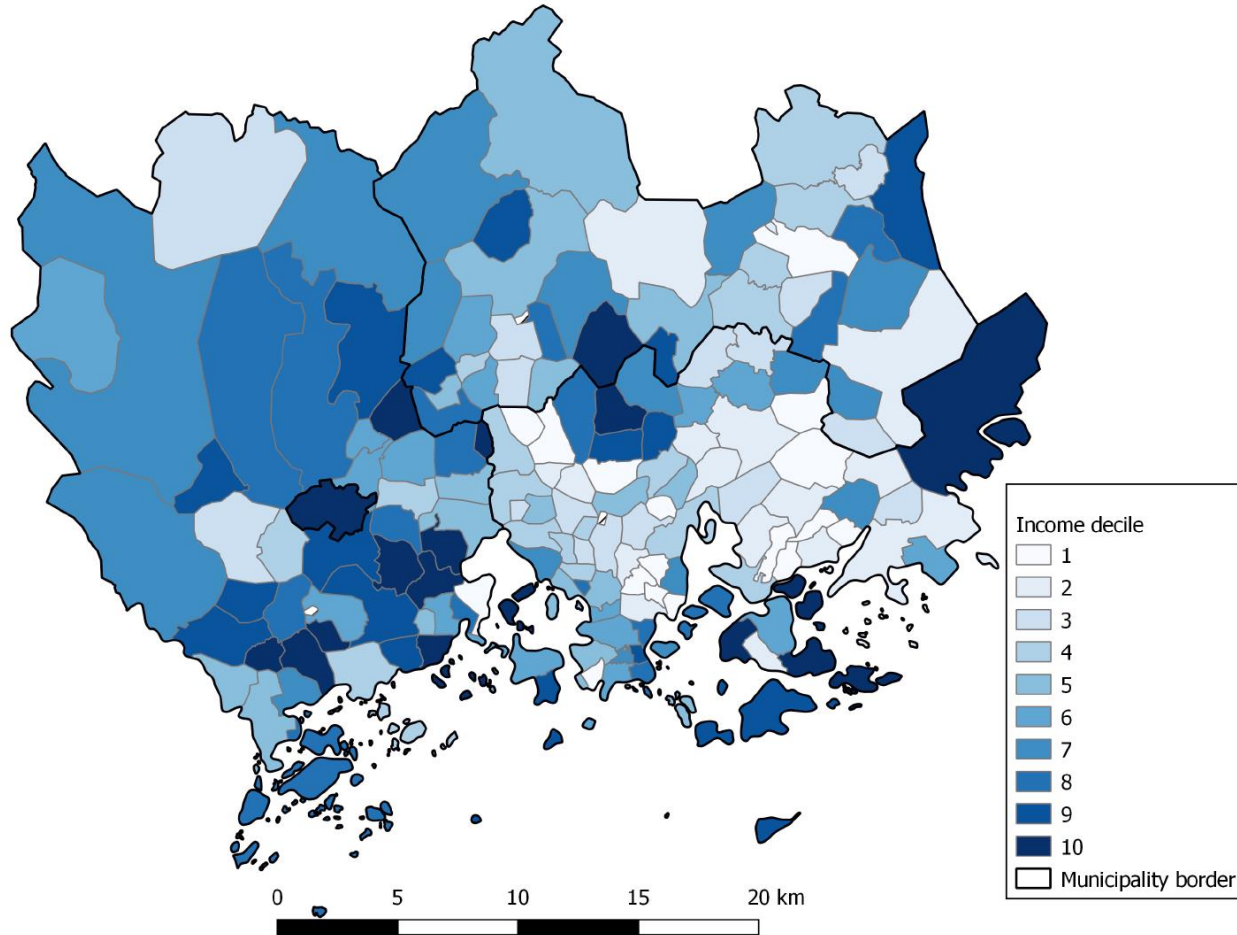
Housing markets and segregation

These three facts lead to a conclusion:

- High-income households can outbid low-income households for houses in the sought-after neighborhoods
- The result is a segregated city by income

Remember that segregation was the outcome also in the monocentric city model when locations differed only w.r.t commuting cost

HMA postcodes

Income decile
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- Municipality border

0   5   10   15   20 km

Housing market mechanism and selection bias

Parents' resources

Location choice: neighborhood quality unrelated to schools

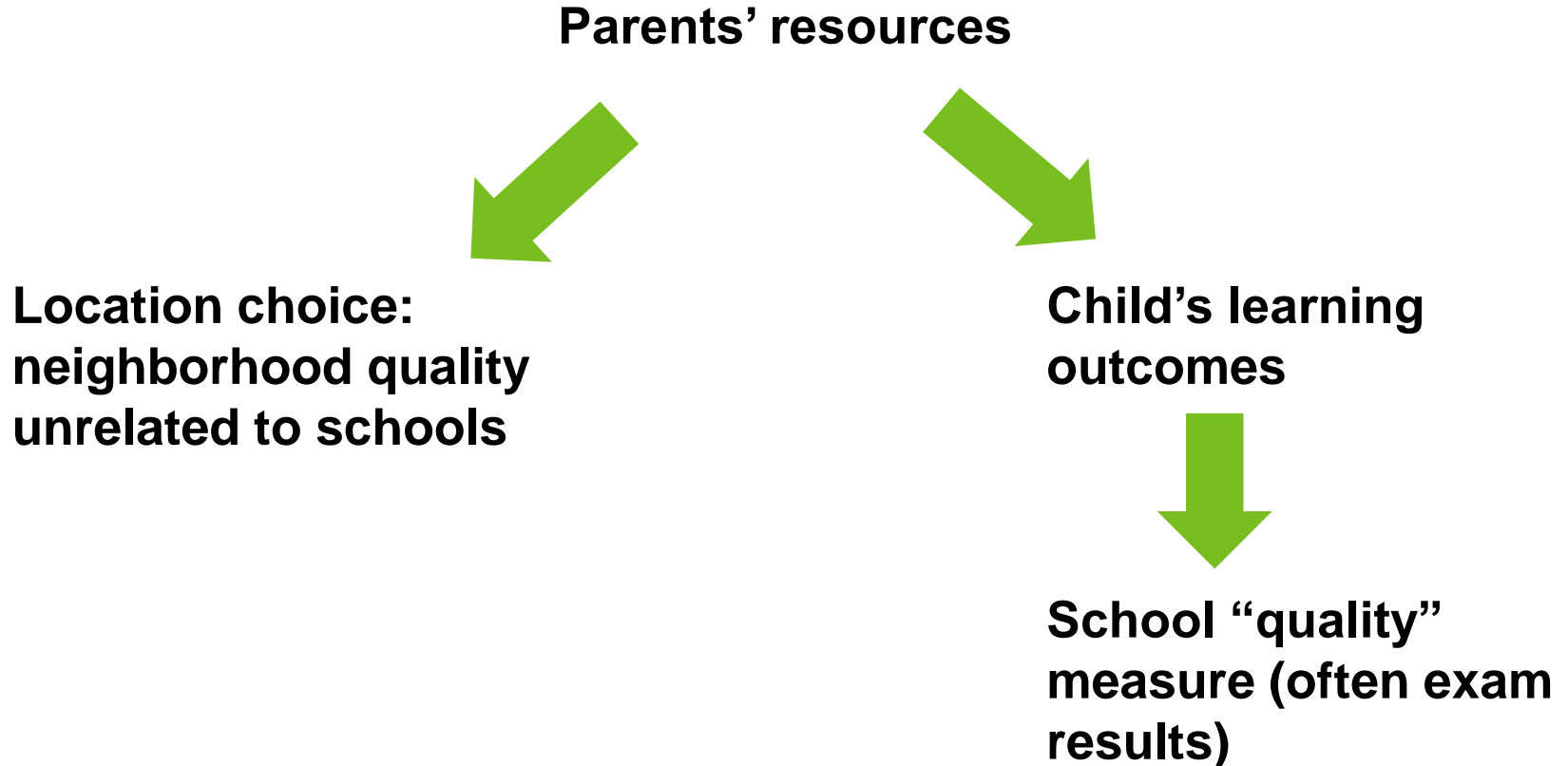# Housing market mechanism and selection bias

Parents' resources

Location choice: neighborhood quality unrelated to schools

Child's learning outcomes

Housing market mechanism and selection bias

Parents' resources

Location choice: neighborhood quality unrelated to schools

Child's learning outcomes

School "quality" measure (often exam results)

Problem in causal inference

The housing market mechanism may lead to a correlation between housing prices and school quality, even if parents do not actually care about school quality

One solution would be to control for observable neighborhood attributes

- The problem is that some n'hood attributes are unobservable (to researcher) and may be correlated with prices and school quality

We need to find a way to plausibly fix all other n'hood attributes and maintain variation in school quality

Boundary discontinuity design

Solution: find areas where school quality varies, but n'hood quality stays fixed

- When access to local public goods is spatially bounded there is a discrete change in space in the quality of the public good

In this case, a solution to this problem is to concentrate on houses at school catchment area boundaries

- Houses near a boundary share the same neighborhood, but the children of the residents are assigned to different schools
- I.e. neighborhood attributes stay fixed, but there is a difference in school quality

School choice in Helsinki

Each housing unit in Helsinki is assigned to an elementary school (*lähikoulu*)

Buying a housing unit within a catchment area secures a place in the school for the child

But children can attend other schools if schools have room (schools don't have to accept)

- Depends on siblings, travel time, aptitude tests and even lottery when schools are full
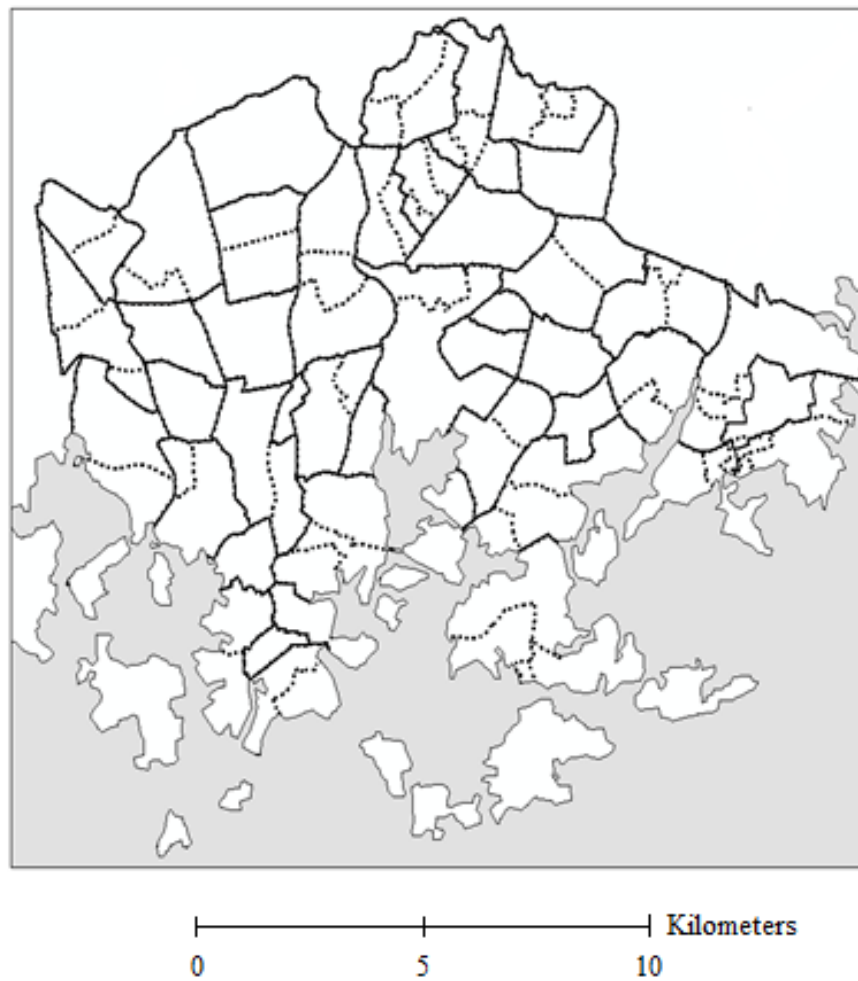
Figure 1. Catchment area boundaries in Helsinki.

Notes: The solid lines represent catchment area boundaries where access to both grades 1–6 and 7–9 changes or boundaries that coincide with geographic obstacles. The dashed lines represent boundaries where access changes only for grades 1–6. The boundaries were obtained from the city of Helsinki.

Boundary discontinuity design
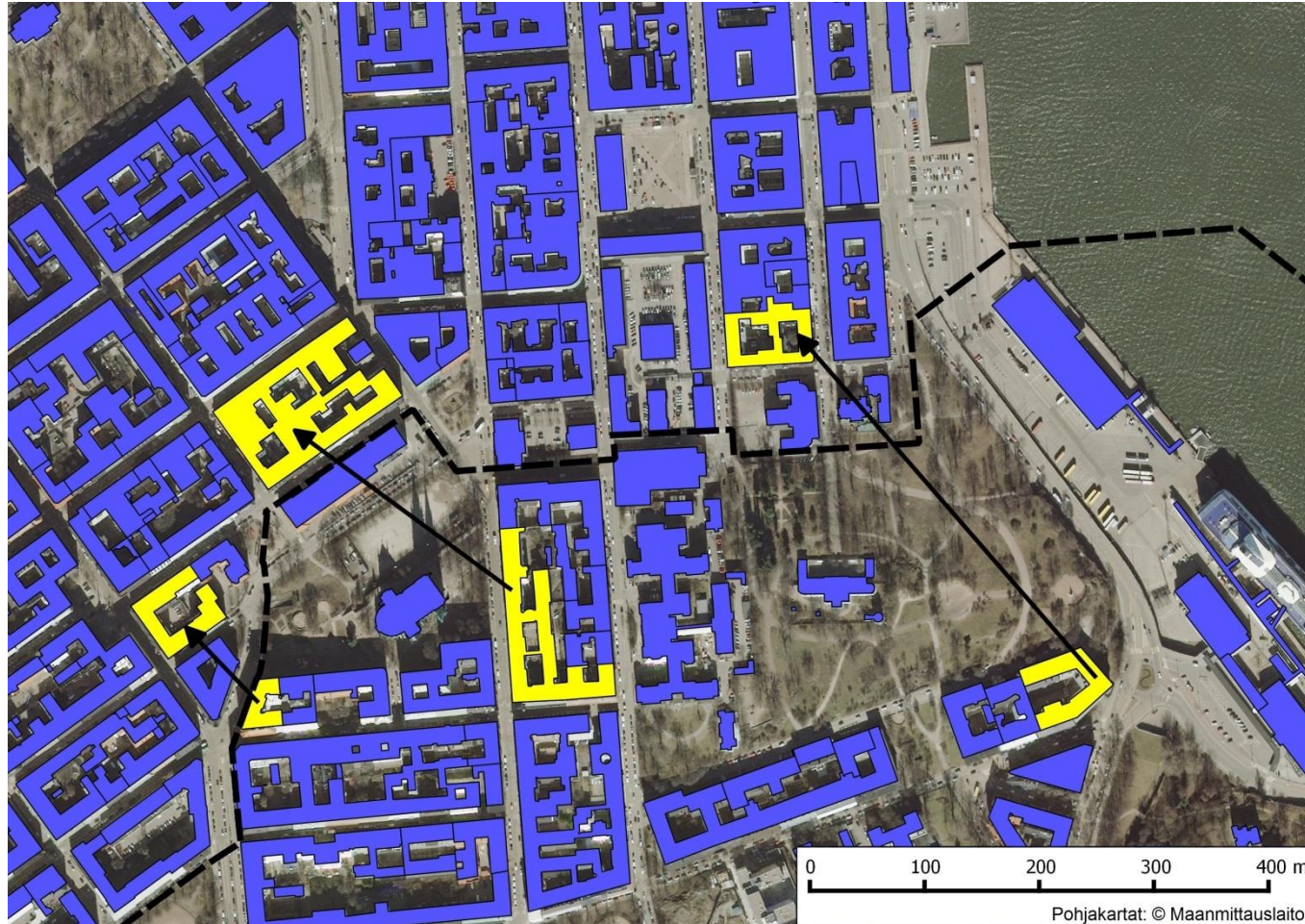
Boundary ----

Transaction

No transaction

29

Spatial differencing

Consider a hedonic model of the form

$$p = \; s(l) \quad \beta + x(l)\gamma + \; g(l) \; + \qquad \varepsilon$$

school quality   house,   unobserved   unobserved, but
at location l n'hood   independent of s, x, g

- A problem arises because of the common dependence of prices, housing attributes and school quality on the unobserved attributes of location l

One solution is spatial differencing:

$$\left(p_i - p_j\right) =$$

$$\left[s(l_i) - s(l_j)\right]\beta + \left[x(l_i) - x(l_j)\right]\gamma + \left[g(l_i) - g(l_j)\right] + \left(\varepsilon_i - \varepsilon_j\right)$$

In practice

For each unit in the data, choose the closest unit on the opposite side of the catchment area boundary

- Same year and same property type

Estimate the hedonic regression model using spatially differenced data with some maximum distance between matches (say, 400m)

- Check robustness also w.r.t maximum distance of matches used in the estimation from 400m to 200m (bias-variance tradeoff)

When does this <u>not</u> work?

Discontinuities in amenities or neighborhood characteristics unrelated to schools

- Geographic features, e.g. major roads or railways
- We can eliminate the most obvious ones

Spatial trends in n'hood characteristics (problem if not enough data exactly at the boundary)

Data

School data from 2008:

- Standardized 6th grade math test scores
- Share of pupils with special needs
- Share of foreign language pupils

Data on individual housing transactions collected by a consortium of Finnish real estate brokers

- Price, unit characteristics, address (years 2008-2012)

Statistics Finland's Grid Database (250m x 250m)

- "Close" n'hood characteristics: household income, education etc.

School characteristics

Table 1. Descriptive statistics for schools characteristics, $N = 50$.

	Mean	Std. Dev.	Min	Max
Math test score	32.1	2.92	21	38
% pupils with special needs	0.09	0.07	0	0.36
% foreign language pupils	0.11	0.10	0	0.44
% pupils going to the school in their catchment area	0.71	0.13	0.36	0.91

	Full sample		Matched sample (< 400 m)	
	Mean	Std. Dev.	Mean	Std. Dev.
Number of observations	14,061		3,852	
Housing unit:				
Price (€)	251,244	152,787	255,211	162,428
Floor area (m^2)	68.6	27.4	67.5	25.1
Age (years)	45.0	33.2	44.1	31.1
Condition (broker estimate):				
Good (0/1)	0.65	0.48	0.65	0.48
Satisfactory (0/1)	0.32	0.47	0.31	0.46
Poor (0/1)	0.03	0.18	0.04	0.18
Building type:				
Row (0/1)	0.09	0.29	0.06	0.25
Multi-story (0/1)	0.91	0.29	0.94	0.25
Own lot (0/1)	0.71	0.45	0.80	0.40
Elevator (0/1)	0.56	0.50	0.61	0.49
Floor level	2.95	1.78	3.10	1.78
Total number of floors	4.57	2.33	4.83	2.22
Maintenance charge (€/m2/month)	3.26	1.07	3.28	1.20
Road distance to CBD (km)	5.94	3.97	6.28	4.55
Distance to match (km)	0.45	0.26	0.23	0.09

|	Full sample		Matched sample (< 400 m)	
	Mean	Std. Dev.	Mean	Std. Dev.
Close neighborhood (250 m x 250 m):				
Homeownership rate	0.51	0.20	0.51	0.19
Mean income (€)	32,231	12,061	32,583	15,183
% college degree adults	0.30	0.12	0.28	0.11
Unemployment rate	0.06	0.04	0.06	0.03
% pension households	0.22	0.10	0.21	0.10
% households with children	0.16	0.10	0.15	0.10
Number of service jobs per capita	0.44	1.20	0.50	1.34
Number of buildings	21.0	12.8	23.5	14.8
Mean floor area of units (m^2)	61.6	17.2	59.3	16.2
Population	734	550	905	673
% foreign language residents	0.09	0.06	0.09	0.05

Main results graphically
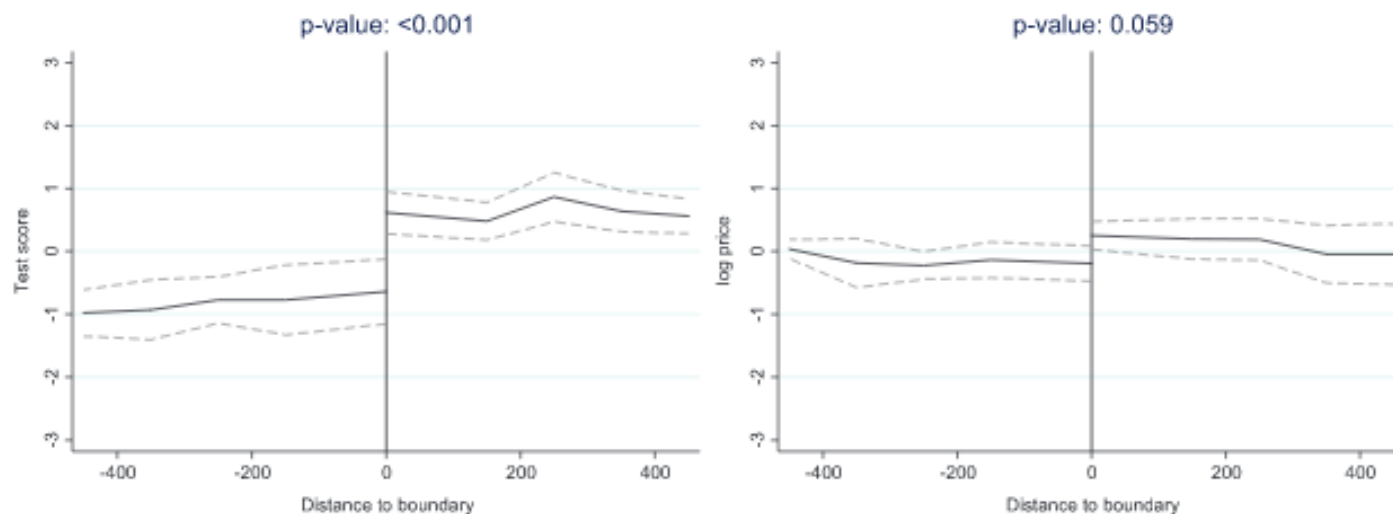
Figure 3. Discontinuities in test scores and log prices.

Notes: Distance to boundary is measured in metres, and negative (positive) distance indicates the side of the boundary with a lower (higher) test score. The *p*-values refer to *F*-tests testing whether the differences are equal at the boundary. The confidence intervals and the *F*-test account for clustering at boundary level.

Main results using regression

Table 3. The effect of school quality measures using cross-boundary differences

	Panel A: Test score			
	(1)	(2)	(3)	(4)
Math test score	0.077**	0.034***	0.036***	0.030***
	(0.031)	(0.010)	(0.008)	(0.006)
	(0.076)	(0.080)	(0.002)	(0.012)
N	3852	3852	3852	3852
R^2	0.09	0.84	0.86	0.87
Unit characteristics	No	Yes	Yes	Yes
Inverse distance weights and boundary distance cubics	No	No	Yes	Yes
Close neighbourhood characteristics	No	No	No	Yes

Main results using regression

	Panel B: Test score and pupil composition			
	(5)	(6)	(7)	(8)
Math test score	0.079**	0.032***	0.041***	0.033***
	(0.037)	(0.009)	(0.013)	(0.011)
	(0.228)	(0.016)	(0.030)	(0.090)
% special needs pupils	0.143	0.077	0.230	0.121
	(0.429)	(0.146)	(0.176)	(0.167)
	(0.830)	(0.707)	(0.424)	(0.633)
% foreign-language pupils	−0.100	−0.183**	−0.166	−0.095
	(0.295)	(0.076)	(0.108)	(0.084)
	(0.729)	(0.058)	(0.214)	(0.358)
N	3852	3852	3852	3852
R^2	0.09	0.84	0.86	0.87
Unit characteristics	No	Yes	Yes	Yes
Inverse distance weights and boundary distance cubics	No	No	Yes	Yes
Close neighbourhood characteristics	No	No	No	Yes

Validity checks

Table 4. Validity checks: varying match distance, excluding zip code boundaries, and using fake boundaries

	(1) Maximum distance <300 m	(2) Maximum distance <200 m	(3) Within same zip-code sample	(4) Fake boundary sample
Math test score	0.043**	0.081***	0.036***	0.004
	(0.018)	(0.024)	(0.011)	(0.008)
	(0.026)	(<0.001)	(<0.001)	(0.713)
% special needs pupils	0.185	0.550*	0.139	0.051
	(0.199)	(0.298)	(0.223)	(0.116)
	(0.517)	(0.112)	(0.639)	(0.783)
% foreign-language pupils	−0.129	−0.018	−0.097	0.024
	(0.114)	(0.127)	(0.086)	(0.090)
	(0.412)	(0.913)	(0.376)	(0.865)
N	2770	1515	2725	4717
R^2	0.87	0.89	0.86	0.85
Unit characteristics	Yes	Yes	Yes	Yes
Inverse distance weights and boundary distance cubics	Yes	Yes	Yes	Yes

Fake boundaries "placebo test"

Figure A3. Catchment area boundaries in Helsinki.

Notes: The solid lines represent the true catchment area boundaries and the dashed lines the fake boundaries.

Conclusions – Harjunen et al. (2018)

A one standard deviation increase in test scores increases prices by roughly 3 percent

- For a two-bedroom apartment this translates to €10,000

Large part of the test score effect can be explained by parents' demand for good peers for their children

At the same time, it seems that parents are not trying to avoid schools with pupils that need extra attention (special needs or foreign language)

Hedonic model and prediction

Using the hedonic model for prediction

In addition to analyzing the price effects of different housing attributes, the hedonic model can used in prediction

For example:

- What would be the value of a given lot for property tax purposes? Mass appraisal of property tax base

- How much cheaper are rent- or price-controlled housing units compared to similar free-market units (in Finland, ARA and Hitas)?

Delivering affordable housing and neighborhood quality: A comparison of place- and tenant-based programs

Essi Eerola, Tuukka Saarimaa*

VATT Institute for Economic Research, P.O. Box 1279, Helsinki FI-00101, Finland

ARTICLE INFO

ABSTRACT

This paper analyzes the relative merits of large place- and tenant-based housing programs in Finland in terms of housing affordability and neighborhood quality. Using hedonic regression methods and household micro data, we find that the rent savings to public housing tenants are considerable and comparable in size to the housing allowance. Furthermore, this public housing subsidy is less targeted towards low-income households than the housing allowance. At the same time, low-income public housing tenants live in poorer, less educated and lower quality neighborhoods than similar low-income households living in private rental housing. This suggests that place-based programs may lead to more segregation than tenant- based alternatives even when neighborhood mixing is an explicit aim of the program, as is the case in Finland.

Analysis in a nutshell

Define subsidy to public housing tenants as

Subsidy = predicted market rent − actual rent

Predict market rents for public housing units using hedonic regression and private market data

- Data on market rents and unit attributes collected from www.vuokraovi.com in 2012 and 2013

- Data on actual rents from the city of Helsinki

Link the estimated subsidy to register data on households

Compare the neighborhoods (zip codes / buildings) of similar low-income hh's in public housing and private rental housing

Predicting rents

Estimate a hedonic rent regression: $\quad p_{ij}^{free} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mu_j + u_{ij}$,

Calculate the subsidy for each public housing unit

$$subsidy_j = \hat{p}_j^{free} - p_j^{sub},$$

The subsidy is correctly estimated only if

- Unobservable unit attributes are not correlated with observable attributes and
- Unobservable unit attributes in the private market are not present in public housing

Otherwise, the difference in predicted and actual rent can arise from omitted unit attributes

Descriptive statistics

Table 1
Descriptive statistics: housing units.

	Private rental market		Public housing	
	Mean	Std. Dev.	Mean	Std. Dev.
Observations	4737		5064	
Rent per m^2 (€)	19.3	4.27	9.81	0.77
Floor area (m^2)	55.5	20.6	59.7	17.5
Number of rooms	2.15	0.87	2.39	0.86
Age (years)	29.0	27.2	32.0	15.8
Balcony (0/1)	0.70	0.46	0.64	0.48
Sauna (0/1)	0.33	0.47	0.04	0.19
Distance to CBD (km)	6.89	3.40	8.36	2.75

Notes: The data on private rental units come from Vuokraovi.com. The public housing rent data come from the city of Helsinki and the public housing unit characteristics data come from Statistics Finland.
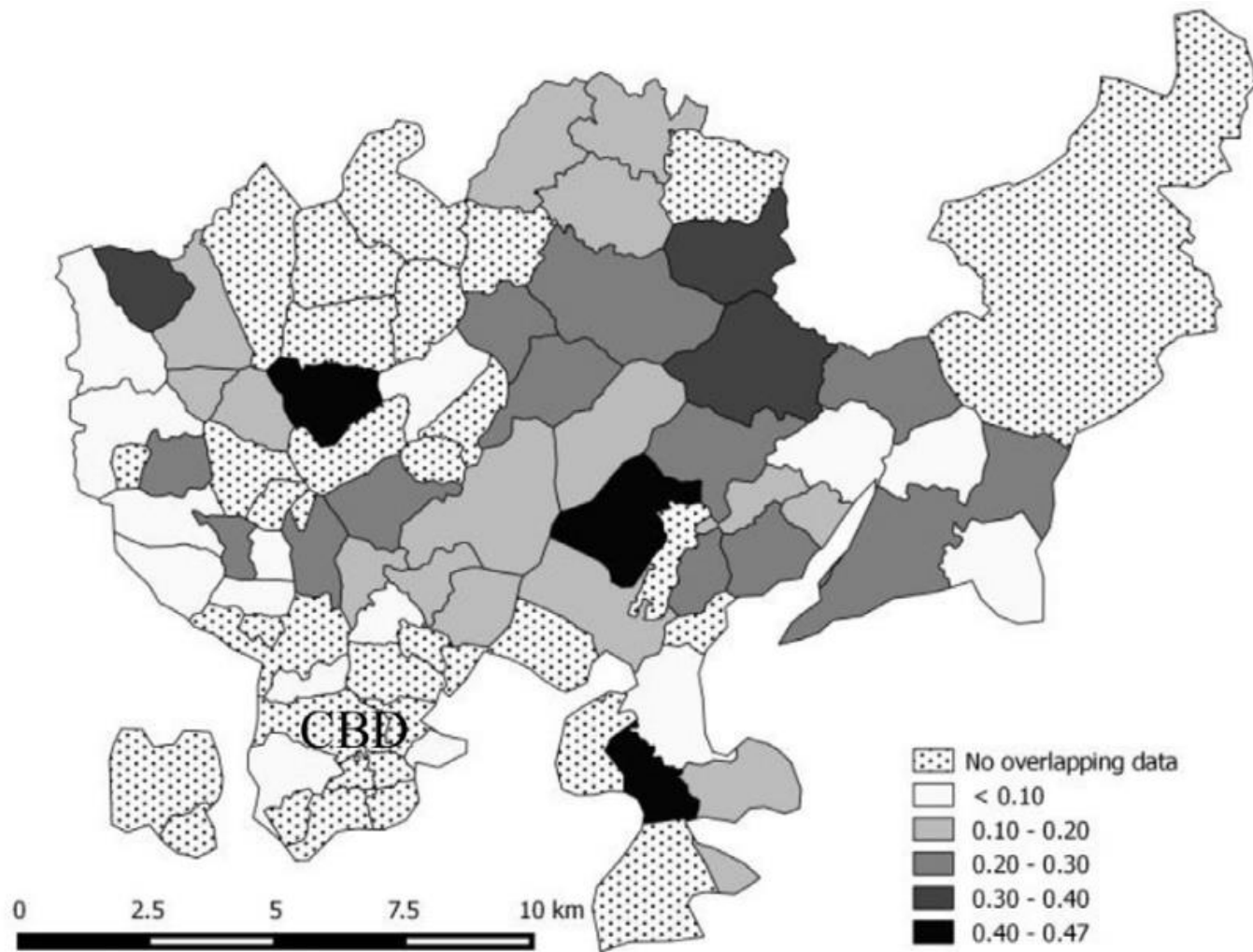
Fig. 1. Share of public housing units by zip code.

Table 2
Hedonic regression results.

	Prediction error for private units (1)	Private units (2)	Public units (3)
Constant	55.10	739.9***	208.5***
	(74.67)	(60.67)	(29.51)
Floor area	4.807	8.987***	9.109***
	(3.723)	(1.601)	(0.646)
$(\text{Floor area})^2$	−0.041	0.011	0.007
	(0.031)	(0.011)	(0.006)
Age	0.603	−4.666***	−5.456***
	(2.282)	(1.075)	(0.906)
Age^2	−0.025	0.044	0.086***
	(0.057)	(0.028)	(0.024)
Age^3	0.00006	−0.00003	−0.0004***
	(0.0003)	(0.0002)	(0.0002)
2 rooms (ref. 1 room)	−42.95*	23.95**	−0.908
	(24.99)	(11.67)	(3.611)
3 rooms	−8.699	60.99***	−4.494
	(31.48)	(13.49)	(5.115)
4 rooms or more	−37.83	89.41***	−12.17
	(58.32)	(24.17)	(9.217)
Sauna (0/1)	−1.115	57.72***	1.251
	(20.37)	(12.49)	(6.734)
Balcony (0/1)	−13.22	−10.08	−3.266
	(26.11)	(10.64)	(2.634)
Distance to CBD	−25.16	−9.650	−2.891
	(24.05)	(8.321)	(7.243)
N	473	4264	5064
R^2	0.17	0.87	0.98

Notes: The table reports results from OLS regressions using housing unit level data. All the models include zip code level fixed effects.
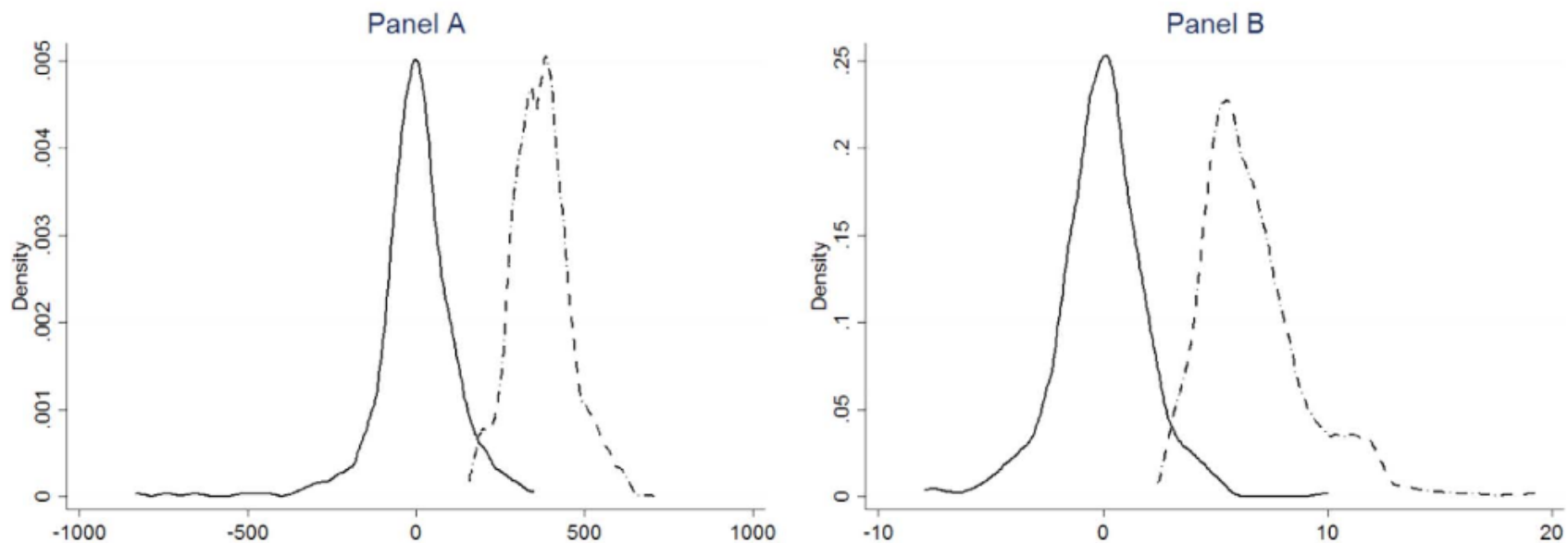
Distribution of the subsidy

Fig. 2. Distributions of public housing subsidy and prediction error for private units.

Notes: Panel A depicts the monthly subsidy (€) and Panel B the monthly subsidy per square meter (€/m²). The solid line refers to private rental units and the dashed line to public housing units.

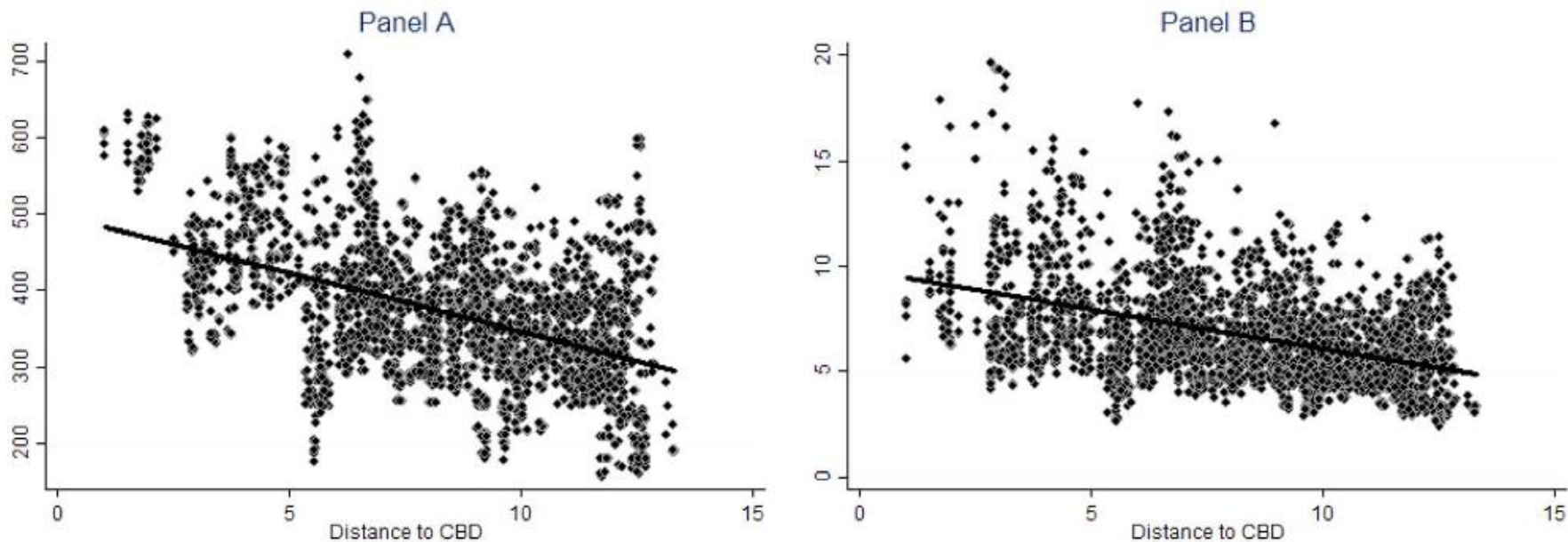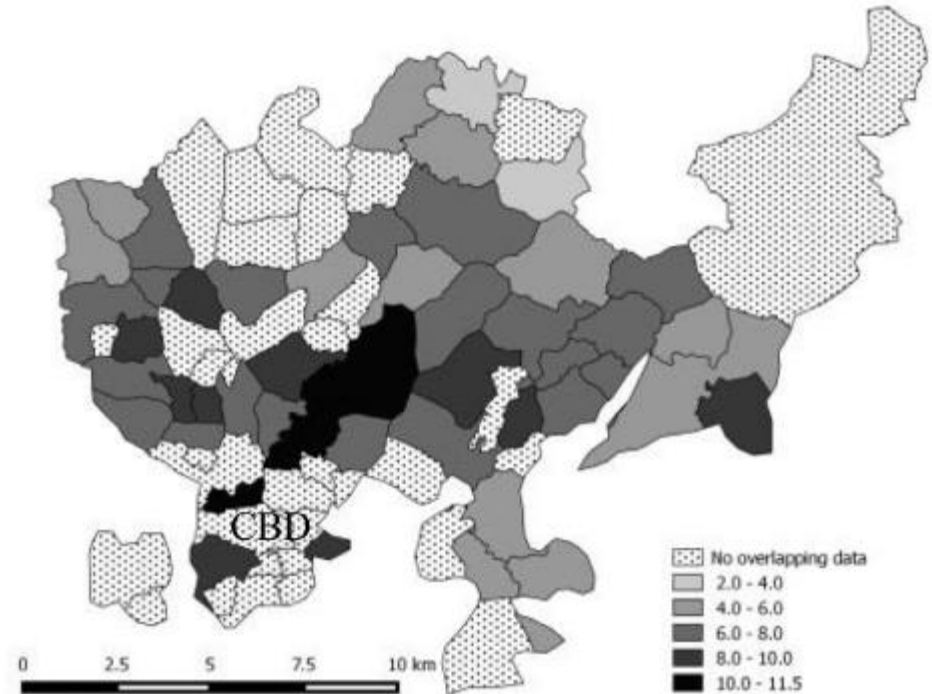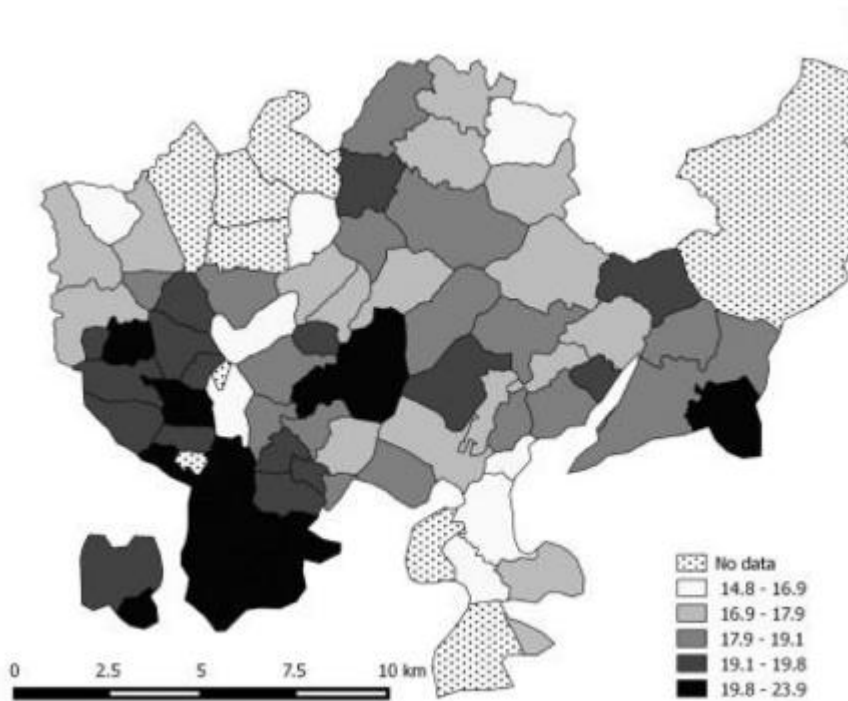# Subsidy according to distance to CBD

Fig. 3. Public housing subsidy according to distance to CBD.

Notes: Panel A depicts the monthly subsidy (€) and Panel B the monthly subsidy per square meter (€/m²).

Market rents and subsidy across neighborhoods

Recap

Housing is a multidimensional product traded in bundles and households value the many features of a house/dwelling

The hedonic model can be used to analyze how housing prices reflect:

- The demand for different housing attributes and
- The demand for public services and neighborhood amenities

The model can also be used for prediction purposes

The model is useful in analyzing the prices and demand of many kinds of differentiated goods