

Saarimaa, Tuukka; Kortelainen, Mika

Conference Paper

Do Homeowners Benefit the Neighborhood? Evidence from Semiparametric Hedonic Regressions

52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Saarimaa, Tuukka; Kortelainen, Mika (2012) : Do Homeowners Benefit the Neighborhood? Evidence from Semiparametric Hedonic Regressions, 52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/120586>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Do Homeowners Benefit the Neighborhood?

Evidence from Semiparametric Hedonic Regressions[†]

Mika Kortelainen^a
Tuukka Saarimaa^b

This version February 2012

WORK IN PROGRESS
DO NOT QUOTE WITHOUT PERMISSION
COMMENTS WELCOME

Abstract

Homeownership is heavily subsidized in many countries mainly through the tax code. The adverse effects of lenient tax treatment of owner-occupied housing on economic efficiency and growth are well documented in the economics literature. The main argument in favor of subsidizing owner-occupied housing is that it creates positive externalities or social benefits that offset these adverse effects. This paper tests whether homeowners create positive externalities to their immediate neighborhood that capitalize into house prices. Using semiparametric hedonic regressions with instrumental variables we find no evidence of positive externalities from neighborhood homeownership rate. This result is robust to relaxing the identification assumptions of our instrument using the set identification method proposed by Nevo and Rosen (2012, *Review of Economics and Statistics*, forthcoming). Our results suggest that the adverse efficiency effects of lenient tax treatment of owner-occupied housing are not offset by positive externalities.

Keywords: Homeownership, neighborhood effects, partial linear model, kernel regression, set identification

JEL codes: D62, R21, R28

[†] We thank James Banks, Rachel Griffith, Christian Hilber, Elias Oikarinen and seminar participants at Helsinki, Jyväskylä, Manchester, Turku and VATT for useful discussions. Oskari Harjunen provided excellent research assistance. Tuukka Saarimaa is grateful to the Academy of Finland and the Finnish Cultural Foundation for financial support and SERC at the London School of Economics and Political Science for hospitality.

^a Government Institute for Economic Research VATT and University of Manchester, E-mail: mika.kortelainen@vatt.fi.

^b Government Institute for Economic Research VATT, E-mail: tuukka.saarimaa@vatt.fi.

1. Introduction

Homeownership is heavily subsidized in many western countries. In most cases, the subsidy is channeled through the tax code by excluding imputed rental income and capital gains from homeowners' taxable income while allowing them to deduct mortgage interest payments.¹ The adverse effects of lenient tax treatment of owner-occupied housing on economic efficiency and growth are well documented in the economics literature.² Furthermore, it has been reported that the tax benefits are regressive and benefit mostly middle- and high-income households.³ Given these facts, the main argument in favor of subsidizing homeownership has to be that it creates positive externalities or social benefits.

The argument for positive externalities from homeownership is based on the hypothesis that homeowners should put more weight on the condition and amenity levels of their neighborhood than renters. This is because in most cases a house is the single most important asset in a homeowner's wealth portfolio, and thus, a homeowner's wealth level depends on the quality of their immediate neighborhood. This should create incentives for homeowners to engage in activities that improve neighborhood quality. If homeowners create social benefits to their neighbors or improve neighborhood quality, neighborhoods with high homeownership rates are more desirable for prospective buyers and higher neighborhood quality should translate into higher house prices.⁴

DiPasquale and Glaeser (1999) and Hoff and Sen (2005) present formal models where homeowners are able to reap the benefits from their investment in neighborhood amenities (broadly speaking) even if they move away from the neighborhood because improved neighborhood quality gets capitalized into house values.⁵ Renters, on the other hand, are unable to capitalize on their investment to the same extent, because their housing costs increase and landlords capture the increased return on housing capital. Moreover, *ex ante*

¹ See Hendershott and White (2000) and Englund (2003) for surveys of different country practices.

² These include inefficient allocation of the capital stock (Berkovec and Fullerton (1992), Skinner (1996) and Gervais (2003)), suboptimal household wealth portfolios (Brueckner (1997) and Flavin and Yamashita (2002)) and frictions in the labor market (Oswald (1999)).

³ See e.g. Hills (1991), Poterba (1992) and Saarimaa (2011).

⁴ It is somewhat unclear whether all these activities should be considered as external benefits in a broader sense. Homeowners may, for example, oppose the building of social housing in their neighborhoods, and thus, only shift the possible harm to other neighborhoods and households. Nevertheless, these activities should raise house prices in the neighborhood.

⁵ Hilber and Meyer (2009) show empirically that house price capitalization induces households to vote for increased school spending.

contracting for these contingencies is impossible. In fact, if rents rise sufficiently improvements in neighborhood quality may even result in a welfare loss for some renters.⁶

Empirical research on the effects of homeownership is growing. A typical empirical strategy has been to look at the effects of homeownership on individual outcomes and behavior, such as voting, civic participation or child achievement. The results from this literature remain inconclusive because of major endogeneity issues in estimation. For example, DiPasquale and Glaeser (1999) find that homeowners are more politically involved. Green and White (1997), Boehm and Schlottmann (1999), Aaronson (2000) and Haurin et al. (2002) find that homeownership is associated with improved child outcomes. However, the most recent studies seem to indicate that some if not all of the positive effects of homeownership found in the earlier literature are driven by inadequate control of unobservable factors that are correlated with homeownership. For example, Barker and Miller (2009) argue that the beneficial effects of homeownership on several measures of child welfare are overestimated in earlier literature. Similarly, using an exogenous social experiment as their identifying assumption, Engelhardt et al. (2010) find that homeownership has no effect on political involvement of households.⁷ Hilber (2010), on the other hand, finds evidence that homeowners are more likely to engage in creating and maintaining social capital. The motivation behind this activity is that social capital is partly capitalized into house prices especially in areas with an inelastic housing supply.

In this paper, instead of looking at the effects of homeownership on outcomes or behavior of individuals, we simply ask: are houses more valuable in neighborhoods with high homeownership rates. Our strategy is to estimate a hedonic house price model where neighborhood homeownership rate is included as an explanatory variable. Coulson et al. (2003a) and Coulson and Li (2011) use this strategy and find a positive association between neighborhood homeownership rate and prices of single-family homes. Both papers use data from the American Housing Survey. A clear problem in using survey data with a hedonic model is that house prices are not from actual transactions but estimated by the owners in the survey. A further possible problem with this data set is the way a neighborhood is defined. It is simply the 10 or approximately 10 housing units nearest to the particular respondent. This

⁶ See e.g. Vigdor (2010).

⁷ Their results are, however, confined to low-income households only and may lack external validity.

type of data does not allow the evaluation of the spatial distribution or the extent of the possible spillovers.⁸

Our analysis differs from and compliments these studies in a number of ways. First, we use geo-referenced house transaction data which facilitates the use of Geographic Information Systems (GIS) and more precise measurement of neighborhood attributes. Second, precise geo-referencing together with semiparametric econometric techniques allow us to control for unobserved neighborhood attributes that do not vary within a relatively small neighborhood.⁹ Second, we concentrate on housing units in multi-story apartment buildings, which are the prevailing housing structure in urban areas. The problem in using data from neighborhoods with single-family houses is that, at least in our application of Finnish data, they are almost exclusively owner-occupied. This makes it problematic to identify the effect of neighborhood homeownership from amenities offered by a neighborhood consisting of single-family houses, such as housing structure, open space, gardens and so forth. Concentrating on apartment buildings in a built-up urban area also practically means that land scarcity and supply elasticity is not a major issue. It is almost impossible or extremely costly to increase density in the study area.¹⁰

Finally, it is likely that any neighborhood effects of homeownership or neighborhood residents are tied to the particular housing structure in the neighborhood. For example, individual homeowners living in single-family houses have a much higher degree of authority on house exterior, yard or garden appearance than their counterparts living in multiunit apartment buildings. Homeowners' incentives may also be tied to housing structure. As Glaeser and Shapiro (2002) and Glaeser (2011) point out the major maintenance problems in multi-unit buildings are building, not unit specific. This leads to an obvious common pool problem as the owners of individual units in a multi-unit building have an incentive to free-ride on the expense of other owners in the building when it comes to investments into common facilities and building attributes. This may apply to neighborhood amenities as well. Linneman (1985) argues along the same lines that particularly in multifamily units and densely populated residential areas landlords may be substantially more efficient in producing housing services compared to individual homeowners. This is manifested, for example, through more efficient maintenance of common facilities and through lower

⁸ For example, Rossi-Hansberg et al. (2010) find a clear spatial pattern in housing externalities related to a neighborhood revitalization program.

⁹ For recent examples of these methods, see McMillen (2010) and McMillen and Redfearn (2010).

¹⁰ Brasington (2002) and Hilber and Mayer (2008) find that capitalization of local amenities is stronger in areas where housing supply is constrained. Hilber (2010) finds that investment in neighborhood social capital is higher among homeowners in built-up neighborhoods.

bargaining costs in the case of externalities or disputes among tenants. It is of obvious interest to see whether homeowners create externalities in an urban environment consisting mostly of multi-story apartment buildings or whether these externalities are confined to single-family housing neighborhoods. These aspects are of interest because in Finland over 40 percent of units in multi-story apartment buildings are owner-occupied.

A caveat in this type of research is that the neighborhood homeownership rate is likely to be endogenous in a hedonic regression model either due to omitted variables or simultaneity. We solve this problem by using instrumental variables. In fact, concentrating the analysis on multiunit buildings offers a natural choice for an instrument. The arguments by Linneman (1985), Glaeser and Shapiro (2002) and Glaeser (2011) suggest that housing structure, and in particular, the number of housing units in a building should drive the homeownership rate of a neighborhood and should be a valid instrument. We elaborate on this below.

We find no evidence of positive externalities from neighborhood homeownership rate that capitalize into house prices. We also test the robustness of these results by relaxing the identification assumptions of our instrument using the method proposed by Nevo and Rosen (2012). In this method, the instrument is allowed to be correlated with the error of the hedonic regression, but the method provides a meaningful set identification result. Using this method we cannot reject the null hypothesis that the true effect of neighborhood homeownership on house prices is zero. These results suggest that the adverse efficiency effects of lenient tax treatment of owner-occupied housing are not offset by positive externalities. This means that the tax favored status that homeownership enjoys in many countries should be questioned.

The rest of the paper is organized as follows. Section 2 introduces the institutional background. In section 3 the econometric model is presented. Section 4 introduces the data and section 5 the results. Section 6 concludes.

2. Institutional background

We utilize data from the city of Helsinki, which is the capital of Finland. The city can be thought of as one housing and labor market region, although commuting from surrounding cities is easy and frequent. The city of Helsinki is an interesting case study also because it is

quite desegregated due strong anti-segregation policies.¹¹ Thus, it is interesting to analyze households' willingness to pay for neighborhood attributes in a city where neighborhoods are more homogenous than probably in most other cities of the same size.

Roughly 45 percent of households living in Helsinki are homeowners. This is a much lower percentage than the national average of over 65 percent. The difference is due to a large number of students living in Helsinki and also because single-family units are more often owner-occupied and single-family units are rare in a dense urban area. In Finland owner-occupied units in multi-unit buildings are part of cooperatives that are incorporated as limited liability companies. This form of ownership is considered as home owning just as much as owning a single-family house and the same tax benefits accrue to both types of home owners.¹² Membership of a co-op is obtained by buying the shares on the open market and no board approval is needed to buy or sell the shares. The shares are treated as private property and can be used as a collateral on mortgage loans just as single family houses. The company owns all the common facilities and usually the lot as well. In some cases, the housing company owns the building but the lot is owned by the city of Helsinki, which leases the lot to the company.¹³ This is an important distinction because neighborhood quality capitalizes into land value and land owners benefit from any improvements in neighborhood quality. Thus, one could argue, that households living on rented lots are not really homeowners in the sense that they are not the residual claimant and do not have incentives to make investments into the neighborhood any more than renter households. However, the contracts are long term (up to 50 years), land rents are well below market rents and renewed infrequently so inhabitants get some, if not all the benefits from capitalization.

Owning a share does not require the owner to live in the unit in question so the owner can freely rent the unit out. In this case, the household living in the unit is registered as a renter. In fact, around half of the privately owned rental units in Finland are rented out this way by private individuals. This institutional setting creates within building variation in homeownership rates across the city.

¹¹ In recent years there has been concern that segregation is increasing in Helsinki, however.

¹² The stamp duty is lower (1.6 percent) when buying housing company shares compared to buying a property (4 percent).

¹³ In fact, the city owns most of the land within its borders. The land owned by the city is mostly recreational (parks etc.) or has subsidized social housing built on it, but also some privately owned buildings and houses are situated on city owned land which the city leases to the inhabitants including single-family homeowners.

3. The data

We use data from two sources. First, house price data are provided to us by the Technical Research Centre of Finland (VTT). The data set is based on transactions where major real estate brokerage firms acted as intermediary. We concentrate on transactions made in Helsinki in 2006 and 2007. The data represent roughly 50 % of all housing transactions in the region during those years. The total sample size is 7554 units.

The second data source is the Grid Database produced by Statistics Finland, which is used to measure neighborhood characteristics.¹⁴ The grid size is 250 m x 250 m and each grid includes information on socio-economic and age structure of the population, building characteristics, employment and service levels. The housing units are geo-referenced to the grids. With this type of data there is some miss-measurement of neighborhood characteristics because it is assumed that the neighborhood characteristics are uniformly distributed within each grid.

Descriptive statistics of the data are reported in Table 1. The housing characteristics included in the data are floor area, age, condition (evaluated by the broker as good, satisfactory or poor), maintenance charge (includes heating, maintenance, property taxes, interest on company debt etc.), indicator that the building is situated on an own or rented lot, elevator, floor level and the number of floors in the building. The data also includes the exact address so we are able to measure exact road distances to the central business district (CBD), to the nearest commuter rail or subway stop and to sea shore using GIS techniques. Table 1 also includes descriptive statistics of the neighborhood variables used in the econometric analysis.

¹⁴ See Statistics Finland (2010) for details of the data.

Table 1. Descriptive statistics.

Variable	Mean	Std. Dev.	Minimum	Maximum
<u>Dwelling characteristics:</u>				
Price (€)	172,545	113,731	60,000	1,500,000
Floor area (m ²)	53.7	25.5	11	362
Age (in years)	54.8	22.6	2	136
Condition:				
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	3.02	1.66	0	9
Total number of floors in the building	4.88	1.74	0	9
Maintenance charge (€/m ² /month)	151.8	77.6	0	1810
Road distances (km):				
CBD	5.80	4.50	0.32	19.16
Nearest train or subway stop	1.25	0.80	0.002	5.83
Sea	1.23	1.32	0.01	7.28
<u>Neighborhood characteristics (grid):</u>				
Homeownership rate	0.49	0.14	0.00	0.98
Personal median income (€)	23,462	4,222	9,150	57,206
Share of adults with a college degree	0.11	0.03	0.00	0.25
Unemployment rate	0.09	0.05	0.00	0.42
Share of pension households	0.21	0.09	0.02	0.81
Share of households with children	0.12	0.07	0.01	0.62
Number of service jobs per capita	0.44	1.00	0.00	31.02
Number of buildings	21	12	2	67
Population	834	568	22	2374
Mean floor area of units	55.5	13.3	30.3	185.8
Number units per building	35.4	18.0	1.1	117.8

Note: The data consist of 7,554 dwelling transactions from Helsinki in 2006 and 2007. All observations are from multi-storey buildings.

4. Econometric model

4.1. Model specification

This section illustrates our empirical approach. As a starting point, consider a hedonic regression model of the partial linear form

$$(1) \quad p_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \delta * amenity_j + f(l_{ij}, la_i) + u_{ij},$$

where p_{ij} denotes the log of transaction price of dwelling i in neighborhood (or grid) j , \mathbf{x}_{ij} is a vector of dwelling and neighborhood characteristics, $f(lo_j, la_i)$ is an unknown function of longitude and latitude coordinates of the dwelling (lo_i, la_i) and u_{ij} is the error term. The function $f(\cdot)$ captures unobservable neighborhood characteristics that do not vary within the local estimation window as explained further below.¹⁵ We assume that homeowners can make investments into neighborhood amenities and social capital, which we denote for simplicity with a scalar *amenity*.¹⁶ According to hedonic theory, the parameter δ can be interpreted as households' mean marginal willingness to pay (MWTP) for a small change in neighborhood amenities.¹⁷ The variables in \mathbf{x} and the coordinates (lo, la) are observed by homebuyers and the econometrician, but *amenity* and u are observed only by the homebuyers, which means that δ is not identified without further assumptions. However, according to theory in DiPasquale and Glaeser (1999), Hoff and Sen (2005) and Hilber (2010) in each neighborhood, the amenity level is a function of the neighborhood's homeownership rate, i.e. $amenity = g(ownrate)$. If homeowners produce positive neighborhood externalities we expect that $\partial amenity / \partial ownrate > 0$.

If we make the simplifying assumption that $amenity = \alpha * ownrate$ and plug this into (1) we get¹⁸

$$(2) \quad p_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \delta\alpha * ownrate_{ij} + f(lo_{ij}, la_{ij}) + u_{ij}.$$

Estimating equation (2) (assuming no endogeneity problems) identifies the product $\delta\alpha$ not the mean MWTP for neighborhood amenities. However, $\delta\alpha$ is the relevant parameter for housing and tax policy purposes because it takes into account both the valuation that homebuyers have for neighborhood amenities and how productive homeowners are in “producing” these amenities. Of course, for a full evaluation of the subsidy to homeowners, we should also know how effective current policies are in encouraging homeownership. This is beyond the scope of this paper.

¹⁵ The model is a variant of local linear models introduced in Cleveland and Devlin (1988). For a recent application of the method in hedonic models, see Rossi-Hansberg et al. (2010).

¹⁶ More precisely, these are not all neighborhood amenities just the ones related to the prevalence of homeowners in the neighborhood.

¹⁷ See Rosen (1974).

¹⁸ The parameter α can be interpreted as “the intensity of neighborhood amenity production by homeowners”. See also Kling et al. (2007).

4.2. Instrument choice

In equation (2), neighborhood homeownership rate is possibly correlated with the error term even with a rich set of controls and the partial linear specification with dwelling coordinates. This could be either due to sorting according to some unobservable neighborhood characteristic that affects prices and attracts homeowners or due to simultaneity of prices and homeownership.¹⁹ Because of possible simultaneity it is also difficult to assess the direction of the bias. Regardless of the reasons for endogeneity an instrumental variable strategy is needed.

Our instrument choice is based on the fact that there are economies of scale in producing housing services that are related to building size in terms of number of units per building.²⁰ This may arise through fixed costs in setting up building maintenance and management. Or that it is cheaper to arrange the cleaning and maintenance of common facilities or tenant selection and monitoring in one building or location compared to many buildings dispersed across space.²¹ Thus, if there are economies of scale, we should expect institutional landlords to own big buildings rather than many small ones for a given amount of investment in real estate. Furthermore, residents in a multi-unit building face a common pool problem because major maintenance problems are building, not unit specific.²² The owners of individual units in a multi-unit building have an incentive to free-ride on the expense of other owners in the building when it comes to investments into common facilities and building maintenance. This may apply to neighborhood amenities as well. However, these common pool problems are not present when a single landlord owns the whole building. Furthermore, landlords may have lower bargaining costs in the case of externalities or disputes among tenants. These arguments suggest that building size and homeownership rate should be negatively correlated.²³

These arguments suggest that housing structure and the number of housing units per building in a neighborhood, in particular, should drive the homeownership rate of a neighborhood. This means that the number of units per building in a neighborhood is a

¹⁹ For example, it could be that user-cost-to-rent ratios vary across neighborhoods due to differences in expected capital gains. See e.g. DiPasquale and Glaeser (1999) and Amior and Halket (2011).

²⁰ See e.g. Glaeser and Shapiro (2002) and Glaeser (2011).

²¹ It could also be that a single landlord (a small investor) would like to own units in the same building due to lower monitoring costs because they know the building, the board that make maintenance decisions etc.

²² See Linneman (1985) and Glaeser and Shapiro (2002) and Glaeser (2011) for details.

²³ Some recent studies also find empirical evidence on this. For example, Hilber (2011) and Lerbs and Oberst (2011) report that housing structure and especially building size is an important driver of whether a housing unit is owner-occupied.

natural choice for an instrument. The identifying assumption is that the number of housing units per building does not have a direct effect on house prices or that it is not correlated with the error term in the hedonic regression model. In order to make this assumption plausible, we control directly for a number of household socioeconomic characteristics along with neighborhood population level and total number of buildings. In addition, as a robustness check we use the method by Nevo and Rosen (2012) that relaxes some of the assumptions of our instrumental variable and still produces a meaningful set identification result.

4.3. Estimation

Following Robinson (1988) the partial linear model can be estimated as follows.²⁴ The first step is to use nonparametric local linear regression to regress p and each variable in the parametric part (\mathbf{x}) individually on the variables in the nonparametric part, in this case the coordinates (lo, la). Then form the residuals from each of these regressions, say e_p and e_{x1}, \dots, e_{xk} .²⁵ The second step is to use OLS to run a regression of e_p on e_{x1}, \dots, e_{xk} . The coefficients for the residuals are consistent estimates of the parameters in the parametric part of the model. When instrumental variables (say \mathbf{z}) are used we simply run the nonparametric regression using the instrument as the dependent variable and coordinates on the right hand side and save the residuals (say e_z). In the second stage, instead of OLS, we run 2SLS using the residuals (e_z) as an instrument for the endogenous variable (or its residual).²⁶

A key issue in the estimation process is choosing the bandwidth or the number of nearest neighbors used in the nonparametric regressions in the first step. A smaller local window (i.e. the number of neighbors) means that the nonparametric part can capture neighborhood quality and unobserved factors at a more local level. However, since our main interest lies on neighborhood amenities we cannot use a very small local window. In this case, a small number of neighbors could result in a situation where some of the neighborhood characteristics would have no or very little variation within the local window. This would make the identification of the impact of neighborhood characteristics infeasible and could also create estimation problems given that in the first step some neighborhood characteristic

²⁴ See also McMillen (2010) and McMillen and Redfearn (2010) for more details.

²⁵ This can be seen as spatial differencing where you subtract the weighted mean value of an observations nearest neighbors from each variable of the observation. Since only coordinates enter the nonparametric part, proximity is simply geographic distance. The mean is basically calculated using inverse distance as weight so that nearer observations get more weight.

²⁶ See Li and Stengos (1996) for details.

are used as dependent variables in a nonparametric regression. Therefore, in this paper we will select the number of nearest neighbors so that there will be enough independent variation for each neighborhood characteristic. We also experiment with bandwidths of different size.

The semiparametric specification mitigates the two problems. First, the model produces a location specific intercept (an estimate of the function $f(\cdot)$), which captures unobservable neighborhood quality that stays constant within the locally fitted regression or local window. This mitigates both endogeneity and spatial autocorrelation problems that may arise due to omitted variables that vary at a broader spatial scale than the local window used in the estimation.²⁷ Thus, the intercept term in the partial linear model works in a similar way as a spatial fixed effect but with one important difference: In the fixed effects specification it is assumed that there is a discontinuous jump at the zip code borders whereas the intercept in the partial linear model varies smoothly over space. Another way to illustrate the difference is to consider a housing unit situated just at a zip code border. In this case, the fixed effects estimator uses only observation within the same zip code in calculating the fixed effects transformation. Obviously, using nearby units from both sides of the zip code border captures unobservable neighborhood quality in a more plausible way. We report results also from a zip code fixed effects model for comparison.

A small local estimation window as possible would be ideal for controlling for neighborhood unobservables. However, in our case the size of the local window or the number of nearest neighbors is constrained from below because the neighborhood variables are measured based on fixed sized grids. If the local window is too small, in some cases all the observations within a local window will be in the same grid and there would be no variation in the neighborhood variables making identification impossible. The average number of transactions in a grid in our data is about 9. However, the maximum number is 143, which is the lower limit for the size of our local window. The average number of observations in a zip code is about 106 whereas the maximum is 446.

²⁷ A similar argument is often made in the school attendance zone boundary fixed effects approach pioneered by Black (1999). They show that including boundary fixed effects and concentrating on observations near attendance zone boundaries controls for unobserved neighborhood quality. Of course, the key difference here is that in a boundary context the discontinuity at the boundary in the key variable is used for identification, whereas we use an instrument that varies within the local window.

5. Econometric results

5.1. Main results

Results from our hedonic models are presented in Table 2. The homeownership rate and other neighborhood characteristics are standardized. As a benchmark we estimated simple OLS models with and without zip-code fixed effects. The OLS model without zip-code fixed effects and controlling for other neighborhood characteristics produces a negative and statistically significant coefficient for neighborhood homeownership rate. At face value, this result means that increasing neighborhood homeownership by one standard deviation (roughly 14 percentage points) decreases house values by approximately 5 percent.²⁸ Including zip-code fixed effects reduces this effect by more than 50 percent to 2 percent.

The partial linear models with small bandwidths (200, 300 and 400 nearest neighbors) produce similar results. In absolute terms, the coefficient is of similar size as in the OLS zip-code fixed effect model. As the bandwidth size is increased the coefficient on homeownership rate increases in absolute terms. This might indicate that smaller bandwidths mitigate omitted variable bias as one would expect.

These models may still suffer from omitted variable bias and especially from reverse causality and an IV strategy is needed. Using 2SLS with zip code fixed effects produces a negative effect, but it is not statistically significant. The first stage F-test statistic is 24.4 indicating that the instrument has good explanatory power. Interestingly, the partial linear models where homeownership rate is instrumented as explained above produce positive coefficients. However, the coefficient is statistically significant only when the largest bandwidth is used. This might be a biased result, because the larger the bandwidth the more likely it is that our instrument is correlated with some unobservable neighborhood characteristic that affects prices. The instrument works well also with the partial linear model.

²⁸ This corresponds roughly to a one standard deviation increase in the homeownership rate. See Table 1.

Table 2. Point estimates from hedonic models.

Panel A: Linear model			
	OLS	OLS	2SLS
Homeownership rate	-0.050*** (0.008)	-0.018*** (0.003)	-0.035 (0.045)
Zip code fixed effects	no	yes	yes
F-test for instrument	24.40
R^2	0.91	0.94	0.93
Observations	7,554	7,554	7,554
Panel B: Partial linear model			
	nn = 200	nn = 300	nn = 400
Homeownership rate (not instrumented)	-0.013*** (0.003)	-0.016*** (0.003)	-0.022*** (0.003)
Homeownership rate (instrumented)	0.032 (0.043)	0.085 (0.052)	0.152** (0.076)
F-test for instrument	22.56	16.73	10.56
Observations	7,554	7,554	7,554

Notes: The dependent variable is the natural log of transaction price. All models include the control variables reported in Table 1 and quarter of sale dummies. Heteroscedasticity robust standard errors are reported in the parentheses. The marking nn refers to the number of nearest neighbors or bandwidth size used in the local regressions in the partial linear model. ***, ** and * indicate 1, 5 and 10 percent significance levels, respectively.

5.2 Robustness check using set identification

The assumption that our instrument does not have a direct effect on prices may not be true. It could be, for example, that large buildings have a direct effect on prices because they block sunlight and view or otherwise make the neighborhood unpleasant. In this section we present results using a method developed by Nevo and Rosen (2012) that relaxes this assumption.

To illustrate the method consider the following simple model

$$(3) \quad p = \mathbf{x}'\boldsymbol{\beta} + \gamma^* \text{ownrate} + u.$$

Let z denote our instrument. Instead of assuming that $\text{corr}(z, u) = \rho_{z,u} = 0$, we need to make further assumptions about the signs of correlations between the error term, the endogenous regressor and our instrument. More specifically, we need to assume that the correlations between the endogenous regressor (*ownrate*) and the error term and between the instrument and the error term have the same sign, i.e. $\rho_{z,u} \rho_{\text{ownrate},u} \geq 0$.

We believe that this is a plausible assumption because of the following arguments. First, due to simultaneity the correlation between the endogenous regressor and the error term is negative, i.e. in neighborhoods with high user-cost-to-rent ratios homeownership rate is lower. A negative correlation would also arise if our instrument in fact is in the hedonic model since we know from the first stage of our 2SLS regressions that the (partial) correlation between the endogenous regressor and instrument is negative. Second, we assume that the correlation between the instrument and the error term is also negative. For example, if sunlight or prevalence of green areas is in the error term a negative correlation arises because neighborhoods with big buildings will have lower levels of these amenities. Moreover, if the instrument is in the hedonic model, it should have a negative coefficient, which also leads to a negative correlation between the instrument and the error term in (3).

The above assumptions provide us with both a lower and an upper bound for the true parameter value, even when the instrument is not exogenous or directly affects prices. The upper bound can be obtained using the standard IV regression or IV with the partial linear model as above.²⁹ The intuition can be easily seen from the probability limit of the standard IV estimator, which can be written as $\text{plim } \beta_{IV} = \beta + \sigma_{z,u} / \sigma_{x,u}$, where σ denotes the covariance. When the second term of the probability limit is positive, clearly the IV estimator gives the upper bound for the true parameter β . Similarly, the lower bound can be obtained using OLS because $\text{plim } \beta_{OLS} = \beta + \sigma_{x,u} / \sigma_x^2$, where σ^2 denotes the variance.

In our case the assumption about the signs of the correlations are enough to identify bounds for the true parameter value. However, tighter bounds can be obtained if we are willing to make a further assumption that the correlation between the endogenous regressor and the error term is larger than the correlation between the instrument and the error term, i.e. $|\rho_{ownrate,u}| \geq |\rho_{z,u}|$. This condition implies that even though we allow $\rho_{z,u} \neq 0$, the instrument needs to be less endogenous than the endogenous regressor (*ownrate*). When this condition is satisfied, one can estimate the lower bound using a generated instrumental variable suggested in Nevo and Rosen (2012). Basically, the generated instrument is constructed as a linear combination of the original instrument and the endogenous variable. Naturally, these correlation assumptions cannot be directly tested because one never observes the error term. However, these assumptions are less strict than the usual assumptions needed for a valid instrument, which assume zero correlation between the instrument and the error term.

²⁹ See Reinhold and Woutersen (2011) for a discussion of the latter case.

The results using the Nevo and Rosen (2012) set identification strategy are presented in Table 3. The bounds are rather large and grow significantly as the bandwidth is increased. However, in each case, zero is in the set of the true parameter value, and thus, we cannot reject the null hypothesis of a zero effect.

Table 3. Set estimates and bootstrapped confidence intervals from hedonic models.

	nn = 200	nn = 300	nn = 400
Upper bound	0.142	0.245	0.829
Lower bound, Nevo-Rosen	-0.021	-0.020	-0.025
Lower bound, not instrumented	-0.021	-0.020	-0.025
No. Obs.	7,554	7,554	7,554

Notes: The table reports 95 percent confidence intervals for the true parameter value estimated using the partial linear model. The confidence intervals are based on 400 repetitions per model.

6. Conclusions

Homeownership is heavily subsidized in many western countries. The adverse effects of lenient tax treatment of owner-occupied housing on economic efficiency and income distribution are well documented in the economics literature. Given these facts, the main argument in favor of subsidizing homeownership has to be that it creates positive externalities or social benefits. If homeowners create social benefits to their neighbors or improve neighborhood quality, neighborhoods with high homeownership rates are more desirable for prospective buyers and higher neighborhood quality should translate into higher house prices.

In this paper, we tested this hypothesis using semiparametric hedonic models with instrumental variables. Our strategy was to estimate a hedonic house price model where neighborhood homeownership rate is included as an explanatory variable. We found no evidence of positive externalities from neighborhood homeownership rate that capitalize into house prices. We also tested the robustness of these results by relaxing the identification assumptions of our instrument using the method proposed by Nevo and Rosen (2012). In this method, the instrument is allowed to be correlated with the error of the hedonic regression, but the method provides a meaningful set identification result. Using this method, we could not reject the null hypothesis that the true effect of neighborhood homeownership rate on house prices is zero.

These results suggest that the adverse efficiency effects of lenient tax treatment of owner-occupied housing are not offset by positive externalities. This means that the tax favored status that homeownership enjoys in many countries should be scrutinized. Possible reforms include taxing the imputed rental income from owner-occupied housing the same way as other capital income, reducing the mortgage interest deduction or making landlords' rental income tax free. The last two suggestions would not result in a neutral tax system, but could be steps toward a more level playing field for homeowners and renters.

References

- Berkovec, J. and D. Fullerton (1992): A General Equilibrium Model of Housing, Taxes, and Portfolio Choice. *Journal of Political Economy* 100(2), pp. 390–429.
- Barker, D. and E. Miller (2009): Homeownership and Child Welfare. *Real Estate Economics* 37(2), pp. 279–303.
- Bajari, P. and M. E. Kahn (2005): Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities. *Journal of Business and Economic Statistics* 23(1), pp. 20–33.
- Bayer P., F. Ferreira & R. McMillan (2007): Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy* 115(4), pp. 588–638.
- Bayer P. & R. McMillan (2008): Distinguishing Racial Preferences in the Housing Market: Theory and Evidence. In A. Baranzini, J. Ramirez, C. Schaefer and P. Thalmann (eds.): *Hedonic Methods in Housing Markets Pricing Environmental Amenities and Segregation*. Springer.
- Black, S. (1999): Do Better Schools Matter? Parental Valuation of Elementary Education. *Quarterly Journal of Economics* 114(2), pp. 577–599.
- Boehm, T.B. & A.M. Schlottmann (1999): Does Home Ownership by Parents Have an Economic Impact on Their Children? *Journal of Housing Economics* 8, pp. 217–232.
- Brasington, D. (2002): Edge versus Center: Finding Common Ground in the Capitalization Debate. *Journal of Urban Economics* 52, pp. 524–541.
- Cleveland, W.S and S.J. Devlin (1988): Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association* 83, pp. 596–610.
- Coulson, N. E., S.E. Hwang and S. Imai (2003a): The Value of Owner-occupation in Neighborhoods. *Journal of Housing Research* 13(2), pp. 153–174.
- Coulson, N. E., S.E. Hwang and S. Imai (2003b): The Benefits of Owner-occupation in Neighborhoods. *Journal of Housing Research* 14(1), pp. 21–48.
- Coulson, N. E. and H. Li (2011): Measuring the External Benefits of Homeownership. Penn State University, Department of Economics, mimeo.
- DiPasquale, D. and E. L. Glaeser (1999): Incentives and Social Capital: Are Homeowners Better Citizens? *Journal of Urban Economics* 45(2), pp. 354–384.
- Engelhardt, G.V., M.D. Eriksen, W.G. Gale and G.B. Mills (2010): What are the Social Benefits of Homeownership? Experimental Evidence for Low-Income Households. *Journal of Urban Economics* 67, pp. 249–258.

- Englund, P. (2003): Taxing Residential Housing Capital. *Urban Studies* 40(5–6), pp. 937–952.
- Gervais, M. (2002): Housing Taxation and Capital Accumulation. *Journal of Monetary Economics* 49(7), pp. 1461–1489.
- Glaeser, E. L. (2011): Rethinking the Federal Bias towards Homeownership. *Cityscape* 13(2), pp. 5–37.
- Glaeser, E. L. and B. Sacerdote (2000): The Social Consequences of Housing. *Journal of Housing Economics* 9(1–2), pp. 1–23.
- Glaeser, E. L. and J. M. Shapiro (2002): The Benefits of the Home Mortgage Interest Deduction. NBER working Paper No. 9284.
- Haurin, D., T.L. Parcel and J.R. Haurin (2002) Does Homeownership Affect Child Outcomes? *Real Estate Economics* 30(4), pp. 635–666.
- Hendershott, P. H. and M. White (2000): The Rise and Fall of Housing's Favored Investment Status. *Journal of Housing Research* 11(2), pp. 257–275.
- Hilber, C. (2010): New Housing Supply and the Dilution of Social Capital. *Journal of Urban Economics* 67(3), pp. 419–437.
- Hilber, C. (2011): The Determinants of Homeownership: Panel Data Evidence from Europe. London School of Economics, mimeo
- Hoff, K. and Sen A. (2005): Homeownership, Community Interactions, and Segregation. *American Economic Review* 95(4), pp. 1167–1189.
- Kling, J.R., J.B. Liebman and L.F. Katz (2007): Experimental Analysis of Neighbourhood Effects. *Econometrica*, Vol. 75, No. 1, pp. 83–119.
- Lerbs O. and C. Oberst (2011): Explaining Spatial Variation in Homeownership Rates: Results for German Regions. CESifo Working Paper No. 3377.
- Li. Q. and T. Stengos (1996): Semiparametric Estimation of Partially Linear Panel Data Models. *Journal of Econometrics* 71(1–2), pp. 389–397.
- Linneman, P. (1985): An Economic Analysis of the Homeownership Decision. *Journal of Urban Economics* 17(2), pp. 230–246.
- McMillen, D.P. (2010): Issues in Spatial Data Analysis. *Journal of Regional Science* 50(1), pp. 119–141.
- McMillen, D.P. and C.L. Redfearn (2010): Estimation and Hypothesis Testing for Nonparametric Hedonic House Price Functions. *Journal of Regional Science* 50(3), pp. 712–733.

- Nevo, A. and A.M. Rosen (2012): Identification with Imperfect Instruments. *Review of Economics and Statistics*, forthcoming.
- Reinhold, S. and T. Woutersen (2011): Endogeneity and Imperfect Instruments: Estimating Bounds for the Effect of Early Childbearing on High School Completion. Mimeo.
- Robinson, P. (1988): Root-N-Consistent Semiparametric Regression. *Econometrica* 56(4), pp. 931–954.
- Rosen, S. (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy* 81(1), pp. 34–55.
- Rossi-Hansberg E., P.-D. Sarte and R. Owens III (2010): Housing Externalities. *Journal of Political Economy*, 118(3), pp. 485–535.
- Saarimaa, T: (2011): Imputed Rental Income, Taxation and Income Distribution in Finland. *Urban Studies* 48(8), pp. 1695–1714. DOI 10.1177/0042098010377474.
- Skinner, J. (1996): The Dynamic Efficiency Cost of Not Taxing Housing. *Journal of Public Economics* 59(3), pp. 397–417.
- Statistics Finland (2010): Grid Database 2010. Available at:
<http://www.stat.fi/tup/ruututietokanta/griddatabase2010.pdf>
- Vigdor, J. (2010): Is Urban Decay Bad? Is Urban Revitalization Bad too? *Journal of Urban Economics* 68(3), pp. 277–289.