

Yan Kestens · Marius Thériault · François Des Rosiers

Heterogeneity in hedonic modelling of house prices: looking at buyers' household profiles

Received: 24 August 2004 / Accepted: 19 July 2005 / Published online: 15 November 2005
© Springer-Verlag 2005

Abstract This paper introduces household-level data into hedonic models in order to measure the heterogeneity of implicit prices regarding household type, age, educational attainment, income, and the previous tenure status of the buyers. Two methods are used for this purpose: a first series of models uses expansion terms, whereas a second series applies Geographically Weighted Regressions. Both methods yield conclusive results, showing that the marginal value given to certain property specifics and location attributes do vary regarding the characteristics of the buyer's household. Particularly, major findings concern the significant effect of income on the location rent as well as the premium paid by highly-educated households in order to fulfil social homogeneity.

Keywords Hedonic modelling · Implicit prices · Heterogeneity · GWR · Household profile

JEL classification R20 · C50

Y. Kestens (✉)

Department of Social and Preventive Medicine, Université de Montréal, 3875, St-Urbain, bureau 328, Montréal, Québec H2W 1V1, Canada

E-mail: yan.kestens@umontreal.ca

Fax: +1-514-4127142

M. Thériault

Centre de Recherche en Aménagement et Développement, Université Laval, 1630-Pavillon Félix-Antoine-Savard, Québec, Québec G1K 7P4, Canada

F. Des Rosiers

Faculty of Business Administration, Université Laval, Québec, Québec G1K 7P4, Canada

1 Introduction

The analysis of house prices using hedonic modelling makes it possible to estimate the marginal monetary contribution of property attributes and neighbourhood externalities (Rosen 1974). In most hedonic models, one unique coefficient is derived for each observed attribute. It is entirely possible that this coefficient may vary according to some systematic pattern. Various methods have been designed to handle such variation (Anselin 1988; Brunsdon et al. 1996; Casetti 1972; Fotheringham et al. 2002; Griffith 1988). Explicitly integrating heterogeneity—which may be spatial—should improve the calibration of the models while enhancing the understanding of the residential market structure.

This paper presents an empirical case study analysing the spatial and social structure of residential property markets by combining single-family property sales and household-level socio-economic data. Through the use of two context-sensitive hedonic methods—the Casetti expansion method (Casetti 1972, 1997) and geographically weighted regression (GWR) (Brunsdon et al. 1996; Fotheringham 2000; Fotheringham et al. 2002)—and through the incorporation of the socio-economic profile of actual property buyers, we have attempted to validate the following hypothesis: the variability of the implicit prices of certain property and location attributes is partly linked to individual preferences. In a recent dissertation, Kestens (2004) showed that the residential choice criteria—both as regards property and neighbourhood—vary significantly with the household profile, that is, with the type of household, age, income, educational attainment, the type of previous tenure (first-time owner vs. former owner), and even with the sense of belonging to the neighbourhood.

In order to investigate these questions, this paper analyses the variation of the impact of property-specifics and neighbourhood attributes considering household socio-economic profiles using hedonic modelling. Thereby, we hope to contribute to Starret's (1981) debate on homogeneity of preferences and capitalisation. As pointed out by Tyrvainen (1997), according to Starret, the capitalisation of an attribute is complete “if: (1) there is enough variation within the variable”—e.g. in order to measure the effect of proximity to power lines, it is important to account for cases where power lines are distant enough so as to prevent any effect on house prices—and “if (2) the residents' preferences are homogeneous. If the preferences are heterogeneous, capitalisation is only partial” (Tyrvainen 1997, p. 220). Whereas the first condition can easily be controlled, the second has been the object of little research. Thus, we hypothesize that the capitalisation is partial in that the value given to an attribute differs with household preferences. While such an assumption may seem to challenge the traditional interpretation of an hedonic function and to question the identification problem addressed by Rosen (1974), it is supported by empirical evidence about the existence of sub-markets and the heterogeneity of hedonic prices over space (Goodman and Thibodeau 2003). We therefore feel that for adequately measuring through hedonic modelling the capitalisation of an attribute, residents' preferences for this attribute have to be homogeneous, or otherwise vary in a

systematic way. In other words, part of the non-stationarity of the value of property and location attributes may be linked to differences among the buyer's household profiles. Appropriate drift-sensitive regression techniques can be used to validate this hypothesis when data is available at the household level. One could argue that standard hedonic models implicitly include buyer's household profiles. Households of similar socio-economic profile select properties in similar locales, the characteristics of which are accounted for by the use of housing, neighbourhood and location variables. However, this paper explicitly tests the marginal effect of property buyer profiles, beyond the average characteristics of the neighbourhood's composition.

The methods presented in this paper should not be considered a valuation tool but merely a way to better understand urban dynamics with respect to house price formation. Results are specific to Quebec City, Canada, and to the socio-economic conditions prevailing in its property market for the 1993–2001 period. Vacancy rates were high and sellers abundant. In this context, advantages in the negotiation process are granted to the buyers, which can explain some of our findings.

Two sets of hedonic models are built using some 761 single-property prices sold in Quebec City between 1993 and 2001. The first set uses Casetti-type interactive terms, while the second relies on GWR. Special attention is given to local indicators of spatial autocorrelation (LISA) (Anselin 1995), as it is expected that the introduction of disaggregated household-level data reduces the number of local spatial autocorrelation “hot spots”. Section 2 discusses the hedonic modelling technique, the spatial dimensions of property markets, and presents Casetti's expansion method as well as the GWR. Section 3 presents the data bank and the modelling procedure, whereas the results are given in Sect. 4. Finally, a summary of the main findings and further research possibilities are presented in Section 5.

2 Literature

2.1 Hedonic modelling

The hedonic framework relies on Lancaster's consumer theory, stating that utility is derived from the properties or characteristics of a good (Lancaster 1966). Since this theory has been extended to the residential market by Rosen (1974), residential hedonic analysis has become widely used as an assessment tool and for property market and urban analysis. The regression of house prices on a variety of property specific and neighbourhood descriptors evaluates their marginal contribution, also called implicit or hedonic prices. In their basic form, hedonic regressions assume each parameter to be fixed in space, which means that each identified attribute has the same intrinsic contribution throughout the submarket under study:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

where \mathbf{y} is a vector of selling prices, \mathbf{X} a matrix of explanatory variables, β a vector of regression coefficients, and ϵ the error term.

However, property markets are very much tied as well as inherent in the spatial structure of the urban landscape. In fact, although capital is mobile, supply may be quite inelastic (Goodman and Thibodeau 1998), and a property, once constructed, becomes immovable, or spatially “rooted”. As a result, the value of a property is largely defined by its location attributes, that is, by its relative location compared with urban infrastructure and services. Furthermore, as pointed out by Goodman and Thibodeau (1998), inelasticities in both supply and demand contribute to market segmentation. As a recent dissertation thesis has shown, the choice criteria concerning both location and property choice vary depending on the household profile (Kestens 2004). This market segmentation may lead to heterogeneous implicit prices, which should be explicitly considered in the residential hedonic price function. In fact, the implicit prices of the hedonic function reflect both supply- and demand-driven forces. In an equilibrium situation, it is assumed that these forces cannot be distinguished within a hedonic function. However, we believe that when the market conditions are not in equilibrium, but instead those of a seller market (much supply for low demand), it becomes possible for the buyers to influence the price they pay for an amenity. If the conditions were reversed, that is, if it were a buyer market (much demand for low supply), the sellers would have more power to impact upon the selling price, and the seller’s characteristics could then be significantly linked to the drift of the implicit prices. Therefore, we assume that the introduction of household-level variables within the hedonic function using appropriate methods like the Casetti expansions may make it possible to estimate the drift in the coefficients associated with certain characteristics of the buyers.

2.2 Spatial dimensions of property markets

Can (1992) distinguishes two types of spatial effects: neighbourhood effects and adjacency effects. The former refers to internalised values of geographical features (exogenous effects), while the latter refers to spatial spill-over effects; that is, the impact of the characteristics of close surrounding properties (endogenous effects). Exogenous effects can be manifold, ranging from city-wide structural factors (e.g. location rent) to local externalities (e.g. view on a high-voltage tower). These geographical features induce trends into housing expenditures that have to be explicitly incorporated into the hedonic function, if they are not removed before modelling.

Classical hedonic modelling would estimate fixed’ coefficients; however, above-mentioned market segmentation may lead to spatial heterogeneity, that is, to possible drifts’ in the estimated coefficients.

Independently from this contextual variation of the impact of housing attributes, similarity of prices between close properties may also be partly linked to spatial spill-over (endogenous effect). Spatial spill-over occurs when characteristics of surrounding or adjacent properties are internalised in the property value, leading to spatial dependency or association. This spatial dependency cannot be modelled adequately using additional descriptive

geographical variables, and necessitates the introduction of spatial autoregressive (SAR) terms into the hedonic function:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{y} + \varepsilon \quad (2)$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \alpha\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \varepsilon \quad (3)$$

where \mathbf{X} is the matrix of explanatory variables, ε the error term, $\mathbf{W}\mathbf{y}$ a spatially lagged dependent variable, with \mathbf{W} as the weight matrix, ρ and α the spatial autoregressive parameters, that is, ρ the degree to which the values at individual locations depend on their neighbouring values, and α the degree to which the values at individual locations depend on their neighbours' residuals (Fotheringham et al. 2002 p. 23).

The SAR terms may take several forms. Most often, however, they are weighted lagged values of the dependent variable (Eq. 2) or of the error term (Eq. 3) (Anselin 1988; Griffith 1988; Kelejian 1995). Ordinary least squares (OLS) is not appropriate for SAR procedures that necessitate generalised least squares (GLS) or Maximum Likelihood (ML) estimations. However, OLS regression presents several advantages: it “has a well-developed theory, and has available a battery of diagnostic statistics that make interpretations easy and straightforward” (Getis and Griffith 2002, p. 131). Spatially dependent variables can also be transformed prior to modelling in their spatial and non spatial components, using spatial filtering techniques (Cliff and Ord 1981; Getis 1995; Getis and Griffith 2002; Griffith 1996, 2000). Of course, combinations of these methods can be used. For example, a model integrating geographical features accounting for the spatial drift may also include an autoregressive term controlling for spatial dependency. However, “a two step procedure is considered to be more suitable” (Can 1990). That means that SAR terms should only be included if spatial dependency is still present after spatial heterogeneity has been fully considered.

2.3 Methods and previous results

In this paper, the spatial heterogeneity of the parameters is handled using two methods, namely, the spatial expansion method developed by Casetti (1972, 1997) and the Geographical Weighted Regression (Brunsdon et al. 1996; Fotheringham et al. 2002). Furthermore, we observe how the introduction of detailed household-profile data helps explaining spatial heterogeneity while diminishing spatial dependence.

The spatial expansion method developed by Casetti has first been used to analyse the spatial drift inherent to various geographical phenomena like migration (Casetti 1986), labor markets (Pandit and Casetti 1989) or price analyses before being applied to property market and price analysis (Aten and Heston 2005, Forthcoming; Can 1990, 1992; Casetti 1997). The parameter drift refers to the variation of the parameter value depending on the context. In fact, this method “extends” fixed parameters by introducing interactive variables that combine a previously defined (fixed) characteristic with a (spatially) dependent variable relating to the (spatial) context:

$$\mathbf{y} = (\mathbf{C}^t(\mathbf{E} + \mathbf{I})\mathbf{X})\beta + \varepsilon \quad (4)$$

with \mathbf{C} , a matrix of contextual variables which can be manifold (including a vector of 1 values in the first column), \mathbf{E} a matrix of expansion indicating which explanatory variables are expanded by the contextual variables, and \mathbf{X} , a matrix of explanatory variables, each one being activated in \mathbf{E} .

In most models' specifications, the estimation of varying parameters is limited to structural factors and the "contextual" variables mainly relate to neighbourhood characteristics [e.g. neighbourhood quality in Can (1990)]. However, the expansion method can be applied more generally, by observing the heterogeneity of any parameter X depending on the "context". This "context" may refer to neighbourhood attributes (quality, distance to the city centre, etc.), but also, as is suggested in this paper, to the specific characteristics of the buyers. The significant expansion parameters therefore measure the variation of the implicit prices people assign to attributes. Also, a parameter can be non significant overall, but may become significant once contextualised. This is only a special case of Eq. (4), that is, when β_0 is null and β_1 is not.

Can (1990) measures the drift of several property specific parameters in relation to the neighbourhood quality for a sample of 577 single-family houses of the Columbus metropolitan area. The two final models consider both the spatial heterogeneity of property specifics (using spatial expansion to neighbourhood quality) and the spatial dependence (using a spatially lagged dependent variable). The parameters that vary significantly through space are the following: the type of exterior, the lot size, the presence of a two-car garage and the presence of a utility room. Recently, a model built with single-family properties transacted during the 1990–1991 period in Quebec City includes several expansion variables (Thériault et al. 2003). Various property attributes are spatially expanded using indicators of relative centrality, family cycle and socio-economic status (derived from census data) as well as using measures of accessibility to regional and local services (computed within a GIS). In addition to age, lot size and connection to the sewer system, three property specifics present spatial drifts: inferior ceiling quality, kitchen cabinets made of hard wood, and the number of washrooms. It seems important to verify whether further drifts in the implicit prices could be related to the buyer's household specific attributes acting on top of the spatial drifts related to social profiles of the neighbourhoods. This research question follows recent findings that showed that the odds-ratio of mentioning a property or neighbourhood choice criteria—i.e., a proxy of their preference for certain types of attributes—is significantly linked to the household profile (Kestens 2004). To the best of our knowledge, no research has yet integrated household profile data into hedonic modelling.

Concomitantly with the expansion method, we ran several GWRs, which gave additional indications on the spatial non-stationarity of the parameters. GWR is an adaptation of moving regressions. Moving regression functions are calibrated for every point of a regular grid, using all data within a certain region around this point. The resulting parameters are site-specific and can therefore vary through space. However, this method is discontinuous, as no weighting schemes are applied to the data used for calibration.

Geographically weighted regressions calibrate local models for every sampling point. However, a weighting scheme (spatial kernel) is applied in order to give greater influence to close data points. Furthermore, the spatial kernel may be fixed (identical for all locations) or adaptive; that is, its bandwidth may vary with the density of the data:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i, \quad (5)$$

where (u_i, v_i) denotes the coordinates of the i th point in space and $\beta_k(u_i, v_i)$ is a realisation of the continuous function $\beta_k(u, v)$ at point i (Fotheringham et al. 2002, p. 52).

Various methods can be used to derive the bandwidth that provides a trade-off between goodness-of fit and degrees of freedom: the generalised cross-validation (GCV) criterion (Craven and Wahba 1979; Loader 1999), the Schwartz Information Criterion (Schwartz 1978) or the akaike information criterion (AIC) (Akaike 1973; Hurvich et al. 1998). For further details on the spatial weighting function calibration, see Fotheringham et al. (2002, p. 59–62). Furthermore, the stationarity of each estimated parameter can be tested using either a Monte Carlo approach (Hope 1968) or the Leung test (See Fotheringham et al. 2002, pp. 92–94; Leung et al. 2000).

In a GWR application on residential value analysis, Brunsdon et al. (1999) showed that the relationship between house price and size varies significantly through space in the town of Deal in south-eastern England.

3 Modelling procedure

All models were built with 761 single-family properties transacted between 1993 and 2001 in Quebec City, Canada (mainly between 1993 and 1996). Property-specific variables were extracted from the valuation role. The characteristics of the vegetation around each property were extracted from remote-sensing data. A Landsat TM-5 image shot in 1999 was categorised using the semi-automated ISODATA (iterative self-organising data analysis) technique, widely used and implemented in some GIS packages (Duda and Hart 1973). Furthermore, the Normalised Difference Vegetation Index (NDVI), a sensitive indicator of the green biomass (Tucker 1979; Tueller 1989; Wu et al. 1997), was derived. For more details about the extraction of vegetation data from remote sensing images and its integration into hedonic models, see Kestens et al. (2004). NDVI is a measure of vegetation density whereas its standard deviation indicates land-use heterogeneity. An additional variable identifies properties with more than 29 trees (according to the number of trees mentioned by the owners during a phone survey, as described below). Previous work by Payne identified this number as the limit upon which the premium accorded to trees was reversed (Payne 1973). Centrality—the mean car-time distance to the main activity centres (MACs)—was computed within a GIS (Thériault et al. 1999). Furthermore, a major phone survey carried out from 2000 to 2003 provided detailed information about each buyer household. The survey concerned the household's moving motivations and property choice criteria, and provided

additional data on the household profiles and on specific attributes of the property, like the number of trees on the lot. A detailed description of the survey and the relations between the motivation to move, choice criteria and household profile are given in Kestens (2004). A description of the variables is presented in Tables 1 and 2.

3.1 Expansion models

In this paper, a first group of models, referred to as global models, is built using the expansion method. All models are in the semi-log functional form (the dependent variable is the logarithm of the selling price) using OLS specification. The first four models (M) omit census variables, whereas the last three (N) include them (see Table 3). A time-drift variable was introduced but did not prove significant. Concerning the M models, a basic model (M1) contains property specifics, vegetation attributes derived from remote-sensing data, and centrality measures, whereas homebuyers' socio-economic variables are added in a second step (M2). Expansion terms (all attributes being "expanded" with regard to the socio-economic profile of the buyers' households) are then added on to both model M1 (resulting in model M3) and model M2 (resulting in model M4). The N series is distinctive in that it contains additional socio-economic Census variables, with N1 as the basic model (including property specifics, vegetation, centrality and Census data), N2 including household profile variables, while expansion terms are introduced in N3. In order to avoid multicollinearity, all expansion terms are built with the previously centered original variables, thereby reflecting the departure from the overall average market values (Jaccard et al. 1990, p. 31).

3.2 GWR models

Concomitantly, using the same dependent and explanatory variables as in M1, M2, N1 and N2, four Geographically Weighted Regressions are built (GWR_M1, GWR_M2, GWR_N1, and GWR_N2). The limitation of the GWR software available, confined to a maximum of 35 variables, made it impossible to derive further GWR versions of models M3, M4 or N3. However, the interest of GWR relies in the possibility of deriving local statistics and a significance test for the stationarity of individual parameters. For a description of further local descriptive statistics that can be obtained using the geographically weighting framework, see Brunson et al. (2002). An *F*-statistic also indicates the significance of improvement between the global and the GWR models. Furthermore, as M3, M4 and N3 are the "expanded" versions of M1, M2 and N2, they can easily be compared with their GWR counterparts, GWR_M1, GWR_M2 and GWR_N2.

All the GWRs were computed with adaptive bi-square spatial kernels, using all data and the AIC minimisation for calibration of the spatial weighting function (Fotheringham et al. 2002, p. 61). The significance test for the heterogeneity of the parameters was made using the Monte Carlo approach (Hope 1968).

Table 1 Description of property specific attributes

Property specifics	Variable name	Description	Type*	Minimum	Maximum	Mean	Standard deviation
	SPRICE	Sale price of the property (Cad \$)	C	53,000	290,000	114,459	40,249
	LNSPRICE	Natural logarithm of the sale price (Cad \$)	C	10.9	12.6	11.6	0.31
	Local tax rate	Local tax rate (/100 of assessed value)	C	1.20	2.76	2.06	0.39
	Living area M2	Living area (sqm)	C	65.7	265.0	122.0	35.3
	Age30-39 × living area	Living area (sqm) × age 30-39	C	-54.9	59.3	-1.8	16.4
	Living area bung.	Living area of a bungalow (sqm)	C	0	206.6	49.7	52.2
	Living area bung	Living area of a bungalow (sqm)	C	-37.7	113.4	-3.6	24.2
	xincome 80 K up	× Yearly income 80,000 Cad \$ up	C	5.3	7.5	6.3	0.3
	LnLotsiz	Natural logarithm of the lot size (sqm)	C	0	54	16.6	12
	App. age	Apparent age (years)	C	-15.6	22	0.1	2.7
	App. age × Single Household	Apparent age (years) × single household	C	0.185	1401.0	143.8	160.1
	Apparent age centreed squared	Apparent age centreed squared	C	-3253.4	2662.2	30.4	501.1
	Apparent age centreed squared × household income	Apparent age centreed squared × household income	C	-1	2	0.00	0.2
	Quality	House quality index	C	0	1	0.56	0.50
	Finished basement	Finished basement	B	0	1	0.49	0.50
	Superior floor quality	Superior floor quality	B	0	1	0.38	0.49
	Facing 51% + brick	More than 50% of facing made of stone or brick	B	0	1	0.14	0.34
	Built-in oven	Built-in oven	B	-0.33	0.54	0.00	0.17
	Built-in oven × age 40-49	Built-in oven × age 40-49	B	0	6	0.34	0.53
	Fireplace	Numbers of fireplaces	C	0	6	-0.03	0.26
	Fireplace first-time owner	Superior floor quality	C	-2.72	0.86	-0.03	0.26
	In-ground pool	Presence of an in-ground pool	B	0	1	0.06	0.23

Table 1 (Contd.)

Variable name	Description	Type*	Minimum	Maximum	Mean	Standard deviation
In-ground pool × household income	Presence of an in-ground pool × household income	C	-4.65	2.90	0.07	0.58
In-ground pool × single household	Presence of an in-ground pool × single household	C	-0.13	0.82	-0.01	0.05
Detached garage	Presence of a detached garage	B	0	1	0.12	0.33
Detached garage × University degree holders	Presence of a detached garage × University degree holder	C	-0.48	0.40	0.00	0.16
Detached garage × couple without child	Presence of a detached garage × couple without children	B	-0.16	0.72	0.00	0.13
Detached garage × Age 40-49	Presence of a detached garage × age 40-49	B	-0.33	0.55	-0.01	0.15
Att garage	Presence of an attached garage	B	0	1	0.08	0.28

*Type of variable: *B* binary, *C* continuous

Table 2 Description of locational and socio-economic attributes

Variable name	Description	Type*	Minimum	Maximum	Mean	Standard deviation
Accessibility						
Car time to MACs	Car time distance to main activity centers (min)	B	4.685	21.0	12.6	3.51
Car Time to MACs Cd Sqd	Car time distance to MAC centered squared (min)	B	1E-06	70.9	12.3	14.7
Car time to MACs centered squared × household income	Car time distance to MAC centered squared (min) × household income	C	-217.2	154.8	3.37	36.3
Highway exit	Car time distance to nearest highway exit (min.)	C	0.04	7.53	2.74	1.46
Household-level data						
Household income	Household income in 10,000 \$ ranges up to 100,000 and up	C	1	10	6.93	2.27
First-time owners	The owner of this property is owner for the first time	C	0	1	0.48	0.50
Age under 30	Aged under 30 at date of transaction	C	0	1	0.04	0.20
Mature trees 100 m	Percentage of area in a 100 m radius covered by residential use with mature trees	C	0	76	18.5	17.20
Mature trees 100 m × Age 30-39	Percentage of area in a 100 m radius covered by residential use with mature trees × age 30-39					
Mature trees 500 m	Percentage of area in a 500 m radius covered by residential use with mature trees	C	0.5	45.0	14.1	9.68
Low tree density 500 m	Percentage of area in a 500 m radius covered by residential use with low tree density	C	2.6	34.3	16.8	5.97
Nb trees 29 up	Number of trees on the property is 29 and up	C	0	1	0.03	0.18

Table 2 (Contd.)

Variable name	Description	Type*	Minimum	Maximum	Mean	Standard deviation
Vegetation						
Nb trees 29 up × age 40 up	Number of trees on the property is 29 and up × age 40 and up	C	-0.31	0.66	0.003	0.09
NDVI standard deviation 1 km	NDVI standard deviation in a 1 km radius	C	0.18	0.60	0.32	0.06
Agricultural land 100 m × University degree holders	(landuse heterogeneity measure) Percentage of agricultural land with dispersed trees in a 100 m radius × University degree holders	C	-21.5	6.68	-0.13	1.63
Agricultural land 100 m × age 30-39	Percentage of agricultural land with dispersed trees in a 100 m radius × age 30-39	C	-9.5	24.2	0.1209	1.7
Agricultural land 100 m × age 40 up	Percentage of agricultural land with dispersed trees in a 100 m radius × age 40 and up	C	-12.5	16.9	-0.1	1.4
Agricultural land 100 m × household with children	Percentage of agricultural land with dispersed trees in a 100 m radius × household with children	C	-11.0	10.0	0.0	1.3
Agricultural land 100 m × first-time owner	Percentage of agricultural land with dispersed trees in a 100 m radius × first-time owner	C	-11.9	20.4	0.07	1.6
NDVI 40 m × age 30-39	Normalised deviation vegetation index within a 40 m radius	C	-0.36	0.32	0.00	0.07
Woodlands 500 m × single household	(density of vegetation) × age 30-39 Percentage of woodlands in a 500 m radius × single households	C	-10.6	26.1	-0.3	3.5
Census data	Percentage of University degree holders	C	4.7	60.1	33.0	14.7
	degree holders in census tract					

Table 2 (Contd.)

Variable name	Description	Type*	Minimum	Maximum	Mean	Standard deviation
Census data						
Percentage of University degree holders × household University holders	Percentage of University degree holders in census tract × household data: University holders	C	-14.8	15.6	2.5	6.9
Percentage of aged 65 up	Percentage of people aged 65 and up in census tract	C	1.3	58.9	8.18	7.33
Developed area (Ha)	Developed area in census tract	C	10.4	407	69.6	32.1
% Dwellings 1946-60	Percentage of dwellings built between 146 and 1,960 in census tract	C	0	62.7	12.3	15.4
Percentage of unemployed aged 15-24	Percentage of unemployed aged between 15 and 24 in census tract	C	0	55.2	18.1	11.4
Nb persons per room	Average number of persons per room in census tract	C	0.3	0.6	0.42	0.06

*Type of variable: *B* binary, *C* continuous

For each model, global and local spatial autocorrelation of the residuals are measured, using Moran's I for the former (Moran 1950) and Getis and Ord's zG^*I (Getis and Ord 1992; Ord and Getis 1995) for the latter.

Table 3 contains the specifications and performance of all models. The estimated parameters, their significance and the Variance Inflation Factor (VIF) values—indicating eventual multicollinearity—are detailed in Table 4 (M series) and Table 5 (N series).

4 Results

4.1 Performance of the global models

Each of the global models explains at least 84% of the house price variation. The best model is N3, with an adjusted R -square of 0.889, a SEE of 10.9%, and an F value of 161. Collinearity is well under control in all models, with only one VIF value slightly exceeding 5 (*Car time to MACs*, model N1).

No model presents any significant global spatial autocorrelation at the 5% level (Moran's I ranges from 0.034 [M4] to 0.172 [M1]). Local autocorrelation is present, but decreases when household-level data is included, and further more when expansion terms are introduced. The number of “hot spots”, that we defined as the significant zG^*i statistics given a 600 m lag (which is the most significant autocorrelation range according to the correlogram), drops from 90 (M1) to 61 (M2) to 41 (M3) to 24 (M4). Results are similar for the N series that includes Census variables: the number of hot spots is already low for N1 (46), and still decreases for N2 (35) and N3 (26). The remaining local spatial autocorrelation in M4 and N3, as defined before, concerns less than 5% of the sample (respectively 24 and 26 cases out of 761, or 3.15% of all cases), and is as such not significant at the 95% confidence level.

The basic models M1 and N1 include classic descriptors as well as several significant variables relating to vegetation, confirming the impact of environmental factors and surrounding land use on house values (Geoghegan et al. 1997; Kestens et al. 2004). The percentage of trees has a global positive impact; however, when the socio-economic condition of the neighbourhood is considered (Census data in Model N1), the impact of vegetation within a 500 m range becomes non-significant. This stresses the links between the socio-economic status of the neighbourhood and land use, mainly with regard to vegetation. Although mature trees in the close surroundings (100 m around the property) represent a premium, the presence of trees becomes detrimental when exceeding a given threshold. In fact, the coefficient for the binary variable identifying properties with more than 29 trees is significantly negative (−5.90%, M1), in accordance with previous findings by Payne (1973).

Accessibility to the MACs is highly significant (t value of −11.02), but the negative effect on property prices is not strictly linear, as proved by the presence of the squared form of the parameter (previously centered to avoid collinearity), with a positive sign (t value 4.41). Hence, the location rent

Table 3 (Contd.)

Global models	Without census data				With census data		
	Model M1	Model M2	Model M3	Model M4	Model N1	Model N2	Model N3
Variables in model			X	X			X
Geographically weighted regression (GWR) models	Interactions (household variables × others)						
	Geographically weighted regression (GWR) models						
GWR models	Model GWR_M1	Model GWR_M2			Model GWR_N1	Model GWR_N2	
Nb of cases	761	761			761	761	
R-square	0.902	0.902			0.885	0.892	
SEE	0.1061	0.1043			0.1098	0.1059	
Number of neighbours used in GWR regressions	321	413			661	707	
F statistic of GWR improvement (sig.)	3.36 (0.002)	3.15 (0.004)			2.78 (0.008)	2.51 (0.013)	
Spatial auto-correlation of residuals	0.045	0.049			0.063	0.082	
	0.364	0.352			0.316	0.265	
Sig.	21	26			26	26	
Nb of significant LISA zG*i statistics (600 m lag, sig. 0.05)	22	21			22	20	
Nb of significant LISA zGi statistics (600 m lag, sig. 0.05)							

Table 4 Coefficients of M models

Dependent variable:	Model M1			Model M2			Model M3			Model M4		
	B	(t value)	Sig	VIF	B	(t value)	Sig	VIF	B	(t value)	Sig	VIF
(Constant)	10.5479	(90.57)	***	10.6027	(97.15)	***	10.8351	(107.78)	***	10.7353	(108.71)	***
Local tax rate	-0.1333	(-9.47)	***	1.5	(-8.47)	***	1.5	(-11.15)	***	1.3	(-10.15)	***
Living area M2	0.0042	(20.6)	***	2.6	(19.94)	***	2.7	(22.35)	***	2.7	(21.16)	***
Age 30-39 × living area	-	-	-	-	-	-	0.0007	(2.98)	***	1.1	-	-
Living area bung.	0.0007	(6.15)	***	2.1	(6.16)	***	2.1	(7.52)	***	2.1	(7.30)	***
Living area bung.	-	-	-	-	-	-	-	-	***	1.1	(-3.12)	***
× income 80 K up	-	-	-	-	-	-	-	-	***	1.1	(-3.12)	***
LnLotsiz	0.1705	(9.17)	***	1.9	(8.54)	***	1.9	(7.59)	***	1.9	(7.93)	***
App. age × single household	-0.0132	(-20.01)	***	3.2	(-19.14)	***	3.0	(-20.24)	***	3.2	(-20.24)	***
App. age centered squared	0.0001	(2.36)	**	1.7	(1.81)	*	0.0035	(2.24)	**	1.0	(2.95)	***
App. age centered squared	-	-	-	-	-	-	1.7	(2.00)	**	1.8	(2.26)	**
× household income	-	-	-	-	-	-	-	-	***	1.3	-	-
Quality	0.0933	(3.67)	***	1.1	(3.65)	***	1.1	(3.27)	***	1.1	(2.68)	***
Finished basement	0.0509	(5.16)	***	1.2	(5.15)	***	1.2	(6.24)	***	1.2	(5.82)	***
Superior floor quality	0.0566	(5.7)	***	1.2	(5.35)	***	1.2	(6.54)	***	1.2	(5.85)	***
Facing 51% + brick	0.0268	(2.76)	***	1.1	(2.08)	**	1.1	(-3.00)	***	1.1	(2.44)	**
Built-in oven	0.0414	(3.04)	***	1.1	(3.39)	***	1.1	(3.66)	***	1.1	(4.33)	***
Built-in oven × age 40-49	-	-	-	-	-	-	-	-	***	1.0	(2.68)	***
Fireplace	0.0259	(2.67)	***	1.3	(2.57)	***	1.3	(3.98)	***	1.4	(3.79)	***
Fireplace × first-time owner	-	-	-	-	-	-	-	-	***	1.1	(2.80)	***
In-ground pool	0.1068	(5.33)	***	1.1	(4.93)	***	1.1	(3.51)	***	1.4	(3.14)	***
In-ground pool	-	-	-	-	-	-	-	-	***	1.7	(2.61)	***
× household income	-	-	-	-	-	-	-	-	***	1.5	-	-
In-ground pool	-	-	-	-	-	-	-	-	***	1.5	-	-
× single household	-	-	-	-	-	-	-	-	***	1.2	(3.12)	***
Detached garage	0.0510	(3.42)	***	1.2	(3.1)	***	1.2	(3.63)	***	1.2	(3.12)	***
Detached garage	-	-	-	-	-	-	-	-	***	1.0	(3.37)	***
× University degree holders	-	-	-	-	-	-	-	-	**	1.0	(-2.42)	**
Detached garage	-	-	-	-	-	-	-	-	***	1.0	(-2.42)	**
× couple without child	-	-	-	-	-	-	-	-	***	1.0	(-2.42)	**

Table 4 (Contd.)

Dependent variable: LnSprice	Model M1			Model M2			Model M3			Model M4		
	B	(t value)	Sig	VIF	B	(t value)	Sig	VIF	B	(t value)	Sig	VIF
Detached garage × age 40-49	-	-	-	-	-	-	-	-	-0.0559	(-1.99)	**	1.0
Attached garage	0.0770	(4.34)	***	1.2	0.0687	(4.08)	***	1.2	0.0682	(4.09)	***	1.2
Car time to MACs	-0.0250	(-11.02)	***	3.2	-0.0271	(-13.1)	***	2.9	-0.0223	(-12.49)	***	2.3
Car time to MACs centered squared	0.0017	(4.41)	***	1.6	0.0018	(5.06)	***	1.6	0.0011	(3.24)	***	1.6
Car time to MACs centered squared × household income	-	-	-	-	-	-	-	-	0.0004	(3.12)	***	1.2
Mature trees 100 m	0.0010	(2.42)	**	2.9	-	-	-	-	0.0007	(1.79)	*	3.0
Mature trees 100 m × age 30-39	-	-	-	-	-	-	-	-	-	-	-	-
Mature trees 500 m	0.0041	(4.44)	***	4.0	0.0036	(5.41)	***	2.4	0.0037	(4.59)	***	3.6
Low tree density 500 m	-0.0034	(-3.32)	***	2.0	-0.0026	(-2.9)	***	1.6	-0.0033	(-3.88)	***	1.6
Nb trees 29 up	-0.0609	(-2.25)	**	1.1	-0.0629	(-2.45)	**	1.1	-0.0689	(-2.71)	***	1.1
Nb trees 29up × age 40 up	-	-	-	-	-	-	-	-	-	-	-	-
NDVI standard deviation 1 km	0.2893	(2.87)	***	2.1	0.2217	(2.46)	**	1.8	-	-	-	-
Household income	-	-	-	-	0.0166	(7.91)	***	1.3	-	-	-	-
First-time owners	-	-	-	-	-0.0427	(-4.72)	***	1.1	-	-	-	-
Age under 30	-	-	-	-	-0.0390	(-1.75)	*	1.0	-	-	-	-
Agricultural land 100 m × University degree holders	-	-	-	-	-	-	-	-	-0.0093	(-3.54)	***	1.1
Agricultural land 100 m × age 30-39	-	-	-	-	-	-	-	-	-0.0091	(-3.50)	***	1.1
Agricultural land 100 m × age 40 up	-	-	-	-	-	-	-	-	-	-	-	-
Agricultural land 100 m × household with children	-	-	-	-	-	-	-	-	0.0098	(3.21)	***	1.1
-	-	-	-	-	-	-	-	-	0.0099	(3.09)	***	1.0

Bold entries: buyers' household characteristics; Italic entries: census variables

***Sig. < 0.01; **Sig. < 0.05; *Sig. < 0.1

Table 5 Coefficients of N models

Dependent variable: LnSprice	Model N1			Model N2			Model N3					
	B	(t value)	Sig	VIF	B	(t value)	Sig	VIF	B	(t value)	Sig	VIF
(Constant)	10.1201	(82.85)	***		10.1710	(86.95)	***		10.2443	(91.49)	***	
Local tax rate	-0.1113	(-7.28)	***	2.0	-0.0963	(-6.54)	***	2.0	-0.0911	(-6.58)	***	2.0
Living area M2	0.0039	(20.18)	***	2.7	0.0036	(19.39)	***	2.8	0.0036	(20.41)	***	2.8
Living area of bungalow	0.0007	(6.23)	***	2.1	0.0006	(6.17)	***	2.1	0.0007	(7)	***	2.1
Ln lot size	0.1800	(10.28)	***	1.9	0.1618	(9.59)	***	1.9	0.1550	(9.61)	***	1.9
Apparent age	-0.0132	(-20.88)	***	3.3	-0.0122	(-19.91)	***	3.4	-0.0113	(-20.01)	***	3.2
App. age centered squared	0.0001	(2.72)	***	1.7	0.0001	(2.39)	**	1.7	-	-		-
Quality	0.0821	(3.44)	***	1.1	0.0767	(3.36)	***	1.1	0.0690	(3.15)	***	1.1
Finished basement	0.0528	(5.69)	***	1.2	0.0489	(5.5)	***	1.2	0.0483	(5.75)	***	1.2
Superior floor quality	0.0528	(5.64)	***	1.2	0.0487	(5.44)	***	1.2	0.0550	(6.59)	***	1.2
Built-in oven	0.0371	(2.89)	***	1.1	0.0415	(3.38)	***	1.1	0.0426	(3.6)	***	1.1
Built-in oven × first-time owner	-	-	-	-	-	-	-	-	-0.0701	(-3.02)	***	1.1
Fireplace	0.0272	(2.99)	***	1.3	0.0262	(3.01)	***	1.3	0.0306	(3.61)	***	1.4
Fireplace × first-time owner	-	-	-	-	-	-	-	-	0.0379	(2.49)	**	1.1
In-ground pool	0.1013	(5.38)	***	1.1	0.0907	(5.04)	***	1.1	0.0746	(4.3)	***	1.1
Detached garage	0.0596	(4.24)	***	1.2	0.0546	(4.07)	***	1.2	0.0513	(3.93)	***	1.2
Detached garage × University degree holders	-	-	-	-	-	-	-	-	0.0799	(3.25)	***	1.1
Detached garage × couple without child	-	-	-	-	-	-	-	-	-0.0790	(-2.5)	**	1.0
Detached garage × age 40-49	-	-	-	-	-	-	-	-	-0.0636	(-2.44)	**	1.1
Attached garage	0.0807	(4.84)	***	1.2	0.0770	(4.82)	***	1.2	0.0763	(4.97)	***	1.2
Car time to MACs	-0.0202	(-7.55)	***	5.0	-0.0215	(-8.4)	***	5.1	-0.0180	(-8.46)	***	3.8
Car time to MACs centered squared	0.0013	(3.45)	***	1.9	0.0014	(3.88)	***	1.9	0.0013	(3.94)	***	1.8
Car time to MACs C. Sq. × household income	-	-	-	-	-	-	-	-	0.0004	(3.76)	***	1.1

Table 5 (Contd.)

Dependent variable: LnSprice	Model N1				Model N2				Model N3			
	B	(t value)	Sig	VIF	B	(t value)	Sig	VIF	B	(t value)	Sig	VIF
Highwext	0.0090	(2.13)	**	2.2	0.0085	(2.1)	**	2.2	-	-	-	-
Mature trees 100 m	0.0014	(4.05)	***	2.1	0.0010	(3.18)	***	2.1	0.0006	(2.02)	**	2.2
Low tree density 500 m	-	-	-	-	-	-	-	-	-0.0018	(-2.58)	***	1.3
Nb trees 29 up	-0.0514	(-2.02)	**	1.1	-0.0475	(-1.96)	**	1.1	-0.0553	(-2.34)	**	1.2
Nb trees 29 up × age 40 up	-	-	-	-	-	-	-	-	-0.1462	(-3.17)	***	1.0
Percentage of University degree holders	0.0050	(9.68)	***	3.4	0.0047	(9.36)	***	3.5	0.0044	(9.17)	***	3.5
<i>Percentage of University degree holders</i> × University degree holders	-	-	-	-	-	-	-	-	0.0018	(3.14)	***	1.1
<i>Percentage of aged 65 up</i>	0.0043	(4.57)	***	2.9	0.0043	(4.7)	***	2.9	0.0038	(4.4)	***	2.9
<i>Developed area (Ha)</i>	-0.0004	(-2.82)	***	1.4	-0.0003	(-2.15)	**	1.4	-0.0003	(-2.26)	**	1.4
<i>Percentage of dwellings 1946-1960</i>	0.0016	(3.31)	***	3.3	0.0013	(2.82)	***	3.3	0.0011	(2.63)	***	3.4
<i>Percentage of unemployed aged 15-24</i>	-0.0013	(-3.31)	***	1.1	-0.0012	(-3.38)	***	1.1	-0.0011	(-3.02)	***	1.2
<i>Nb persons per room</i>	0.4368	(3.72)	***	3.0	0.4574	(4.07)	***	3.0	0.3692	(3.37)	***	3.1
Household income	-	-	-	-	0.0145	(7.23)	***	1.3	0.0155	(7.99)	***	1.3
First-time owners	-	-	-	-	-0.0396	(-4.68)	***	1.1	-0.0417	(-5.1)	***	1.1
NDVI 40 m × Age 30-39	-	-	-	-	-	-	-	-	-0.2110	(-3.45)	***	1.1
Woodlands 500 m × single household	-	-	-	-	-	-	-	-	-0.0027	(-2.49)	**	1.0
Agricultural land 100 m	-	-	-	-	-	-	-	-	-	-	-	-
× University degree holders	-	-	-	-	-	-	-	-	-0.0083	(-3.42)	***	1.1
Agricultural land 100 m	-	-	-	-	-	-	-	-	-	-	-	-
× first-time owner	-	-	-	-	-	-	-	-	-0.0054	(-2.14)	**	1.1

Bold entries: buyers' household characteristics; Italic entries: census variables

***Sig. < 0.01; **Sig. < 0.05; *Sig. < 0.1

follows a quadratic function and takes the form of a U-shaped curve, with positive premiums both in the city centre and in the outer suburbs, *ceteris paribus*. A previous study showed that land-use and vegetation attributes significantly explain part of these premiums, reducing the value and significance of the squared distance term (Kestens et al. 2004). Therefore, if vegetation descriptors were absent, this parameter would be even higher and more significant.

4.2 Introduction of socio-economic variables describing the household

Three variables describing the household are significant : the household income and the previous tenure status (Models M2, M4, N2 and N3) as well as the age of the respondent at transaction date (under 30) (Model M2 only). *Ceteris paribus*,

- For each additional \$10,000 of income, buyers pay an average premium of 1.61% (1.46–1.73%, depending on the model)
- First-time owners pay between 4.04 to 4.36% less than former owners
- Young households, under 30 years of age, pay 3.98% less than older buyers for the same property (only model M2, and sig 0.1).

Whether Census variables—describing the socio-economic profile of the neighbourhood at the Census-tract level—are included or not in the model, the two household-level variables Household Income and First-time Owners stay significant, with similar and high t values (ranging from 7.23 to 8.47 and from 4.68 to 5.1, respectively, depending on the model). Furthermore, no significant collinearity is detected between the two levels of socio-economic measures (Census data and household data), the maximum VIF value among these variables being 3.5 (Percentage of university degree holders in the Census tract, model N2).

Concerning the dichotomous age variable (Under 30), it is present in one model only (M2), with a low significance test (t value -1.75 , sig. 0.1). Although it does not present any collinearity with household income or previous tenure status as could have been expected, this variable drops out when Census data (N2) or further expansion terms are included (M3, M4, N3).

4.3 Adding expansion terms: controlling for heterogeneity

In a last step, we introduced expansion terms allowing for the basic parameters [property specifics, accessibility, vegetation (M3 and M4) and Census data (N3)] to vary with regard to the household profile. Several expansion terms are significant, showing that the value given to certain property specifics or location attributes is not homogeneous among buyers. Table 6 presents the list of the parameters that are heterogeneous considering the household characteristics of the buyers.

While a majority of expansion terms (15) is significant when both Census data and raw household profile variables are omitted (Model M3), only a few

Table 6 Synoptic table of significant expansion terms

The value given to the...	...varies regarding the buyer's...	Age	Income	Household type	Educational attainment	Previous tenure status
Property specifics	Living area	X				
	Living area of a bungalow		X			
	Apparent age		X	X		
	Built-in oven	X				X
	Fireplace					X
	In-ground pool		X	X		
	Detached garage	X		X	X	
Centrality	Car-time to MACs	X				
Vegetation	Mature trees 100 m	X				
	Nb trees 29up	X				
	Agricultural land 100 m	X		X	X	X
	Woodlands 500 m			X		
	NDVI 40 m (greenness)	X				
Neighbourhood profile (census)	Percentage of University degree holders				X	

Nb: The significant buyer's household variables may vary depending on the interaction considered. For example, Age may refer to several categories (age 30–39, age 40–49, age 40 up, etc.). See Tables 4 and 5 for complete details.

drop out when these are included (12 interactive variables in both models M3 and N3). Also, some parameters are only significant when their non-stationarity is considered, as NDVI 40 m, Woodlands 500 m and Agricultural Land 100 m. These variables are not significant as such but need to be expanded to enter the model. This shows that for some attributes, estimating a unique coefficient for the whole area of study is not possible, and that the spatial variability must be considered in order to properly measure their impact.

4.4 GWR models

The variables of the four models M1, M2, N1 and N2 were introduced in four GWRs, resulting in GWR_M1, GWR_M2, GWR_N1 and GWR_N2. These models performed well, with *R*-squares ranging from 0.885 to 0.902 (see Table 3). The *F*-statistics of improvement between global and GWR models, however low (values ranging from 2.51 to 3.36), are significant.

As expected, no global autocorrelation is left in the models. Some local “hot spots” are still significant here too, but represent less than 5% of the sample (21–26 significant zG^*i statistics for a spatial lag of 600 m.). Figure 1 shows a map of significant zG^*i statistics for GWR_N2.

Geographically weighted regression gives the possibility of deriving local regression statistics, for example the local significance of a parameter. As GWR calculates distinct regressions for each point of the sample, the variability of the significance can be mapped. Furthermore, the non-stationarity can be tested using a Monte Carlo approach. That is, the question is to know

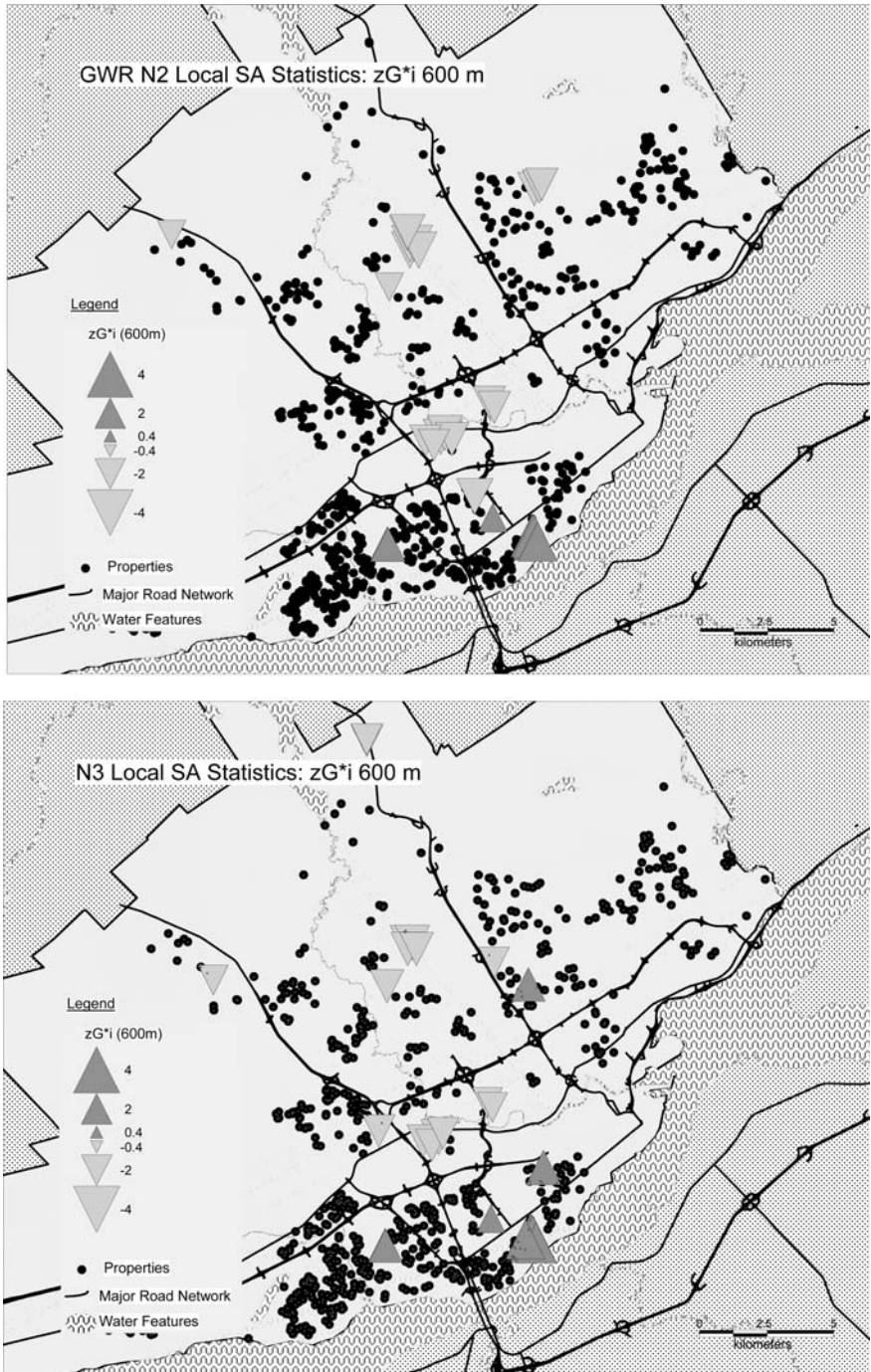


Fig. 1 Local spatial autocorrelation: significant zG*i statistics for N3 and geographically weighted regression (GWR)_N2

whether the observed variation is sufficient to say that the parameter is not globally fixed. p values testing for non-stationarity are given in Table 7. For the parameters with non-significant p values, it is assumed that a unique coefficient holds true. The parameters that are considered non-stationary are therefore the following: local tax rate, apparent age (Fig. 2), Car Time to MACs (Fig. 3), NDVI Stdd. 1 km (GWR_M1 and GWR_M2), and % University Degree Holders (GWR_N1). Also, local R -squares give further indication about the fit of the model depending on location. However, the value of the local R -square is also influenced by the stationarity of the

Table 7 Non-stationarity of parameters in GWR Models (p values) and Moran's I statistic

Parameter	p value				Moran's I (1,500 m)
	GWR-M1	GWR-M2	GWR-N1	GWR-N2	
(Constant)	0.02	0.06	0.39	0.32	
Local tax rate	0.00	0.00	0.08	0.10	0.66
Living area m ²	0.10	0.17	0.58	0.53	0.47
Living area bung.	0.27	0.39	0.56	0.79	0.53
Ln lot size	0.07	0.10	0.16	0.11	0.94
App. age	0.01	0.02	0.27	0.40	1.10
App. age centered squared	0.50	0.49	0.60	0.77	0.98
Quality	0.24	0.32	0.49	0.38	0.04
Finished basement	0.26	0.17	0.35	0.34	0.08
Superior floor quality	0.32	0.30	0.25	0.38	0.31
Facing 51%+ brick	0.50	0.24	–	–	0.44
Built-in oven	0.15	0.04	0.38	0.37	0.08
Fireplace	0.64	0.76	0.36	0.20	0.29
In-ground pool	0.88	0.90	0.45	0.62	0.07
Detached garage	0.38	0.48	0.32	0.59	0.29
Attached garage	0.90	0.67	0.71	0.40	0.02
Car time to MACs	0.00	0.02	0.17	0.32	0.78
Car time to MACs centered squared	0.00	0.00	0.13	0.21	0.96
Highway exit	–	–	0.09	0.22	0.81
Mature trees 100 m	0.75	–	0.37	0.28	0.88
Mature trees 500 m	0.49	0.32	–	–	1.09
Low tree density 500 m	0.15	0.08	–	–	0.60
Nb trees 29 up	0.58	0.58	0.41	0.28	0.02
NDVI Standard deviation 1 km	0.00	0.00	–	–	0.95
Household income	–	0.45	–	0.41	0.27
First-time owners	–	0.73	–	0.97	–0.11
Age under 30	–	0.41	–	–	–0.05
<i>Percentage of University degree holders</i>	–	–	0.01	0.13	0.80
<i>Percentage of aged 65 up</i>	–	–	0.89	0.46	0.75
<i>Developped area (Ha)</i>	–	–	0.11	0.17	0.81
<i>Percentage of dwellings 1946–1960</i>	–	–	0.87	0.90	0.90
<i>Percentage of unemployed aged 15–24</i>	–	–	0.22	0.28	1.30
<i>Nb persons per room</i>	–	–	0.66	0.60	0.59

Italic entries: census data variables; Bold: buyers' household variables
 Bold numbers: significant at the 95% confidence level

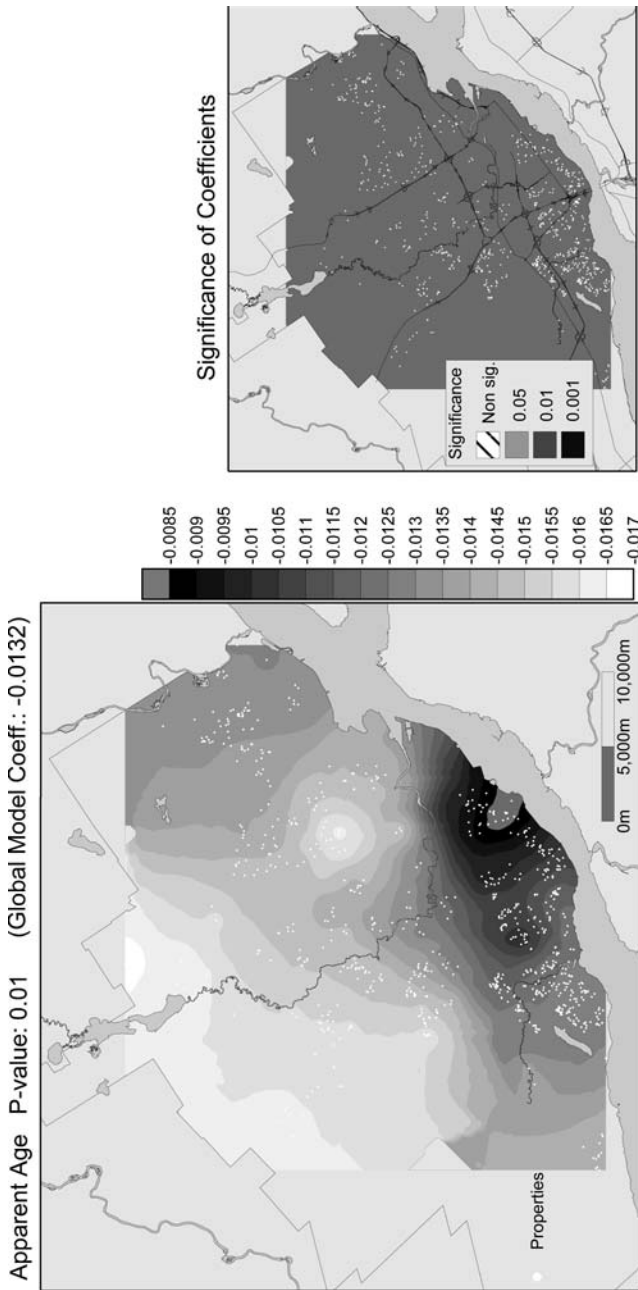


Fig. 2 Geographically weighted regression_M1: spatial variation of apparent age parameter

M1 GWR Parameters Mapping

CBD Car Time P-value: 0.00 (Global Model Coeff.: -0.025)

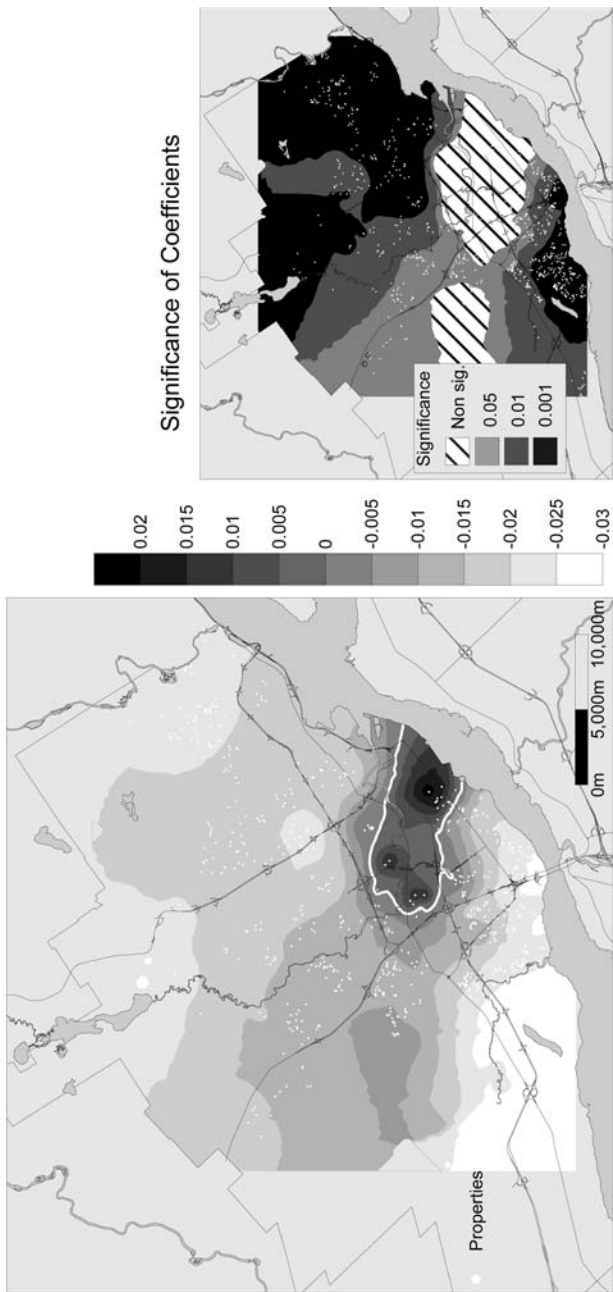


Fig. 3 Geographically weighted regression_M1: spatial variation of car-time to MACs coefficients NB: the non-significance of the coefficients in certain areas is partly due to the scarce presence of single-family properties, as for example in the more central positive-sign area

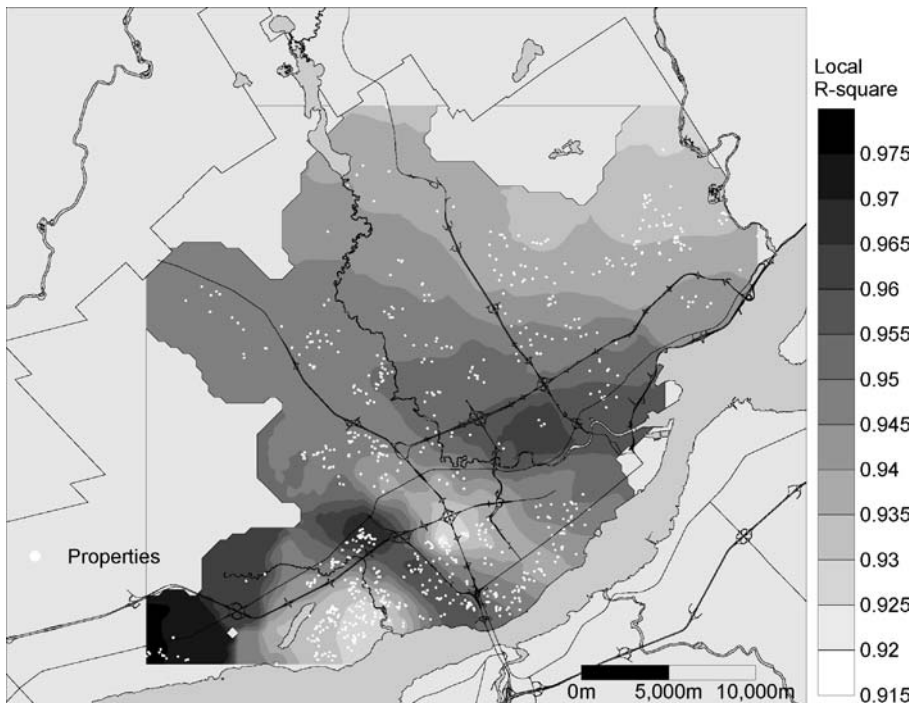


Fig. 4 Local R-squares for Geographically weighted regression_M2

process that is modelled. Therefore, this statistic should be interpreted with care (See Figure 4 as an example of a map of local R-squares for GWR_M2).

4.5 A comparison of global and GWR models

Although the GWR models must be compared with their global counterpart (that is, with the global models built with the same variables, *M1*, *M2*, *N1*

Table 8 Spatial structure of the household characteristics that explains the heterogeneity of parameters (expansion models)

	Moran's I (1,500 m)	Probability value
Household income	0.271	0.02
Income 80 K up	0.231	0.04
Age 30–39	−0.109	0.21
Age 40–49	−0.158	0.12
Age 40 up	−0.022	0.44
Household with children	−0.219	0.05
Couple without child	−0.053	0.35
Single household	0.145	0.14
University degree holders	0.360	0.00
First-time owner	−0.105	0.22

Bold: significant at the 5% level

and $N2$), it is also of interest to compare the GWRs with the expanded versions of the global specifications. For example, let us compare the two “drift”-sensitive versions of $N2$, that is $N3$ and GWR_N2 . In both cases, the percentage of explanation of the variance is similar (0.894 for the global version, vs. 0.892 for the GWR), as is the global autocorrelation of the residuals (Moran’s I values respectively 0.102 and 0.0802). Concerning the local autocorrelation, the number of significant zG^*i statistics (26) is identical, although these hot spots do not strictly match spatially (See Fig. 1). In the end, these models are similar in terms of explanation power and for their ability to handle spatial autocorrelation.

Let us compare more precisely how these models handle heterogeneity. For $N3$, the coefficients that vary spatially are identified by the significant expansion terms. These expansions refer to the following variables: built-in oven, fireplace, detached garage, car time to MACs, Nb of trees 29 up, percentage of University degree holders, NDVI 40 m (greenness) Woodlands 500 m and agricultural land 100 m. The statistical significance of expansion terms indicates that for these variables, a single coefficient is not a valid alternative. In fact, we know that the impact of these variables varies according to age, income, educational attainment and type of household. However, no local measure of significance is available.

For the GWRs, the heterogeneity of the parameters is given by the p values measured through a Monte Carlo procedure (Table 7). According to these p values, four parameters vary significantly at the 95% confidence level for GWR_M1 and GWR_M2 [local tax rate, apparent age, car time to MACs (linear and squared form) and NDVI Stdd. 1 km (heterogeneity of land use), one for GWR_N1 (percentage of university degree holders), and none for GWR_N2 . It is interesting to note that each of these variables identified as non-stationary is also strongly spatially structured, as indicate the corresponding high Moran’s I statistics (Table 7, fourth column). Also, the findings suggest that for the variables with non-significant p values, a unique coefficient is adapted, that is, the implicit price is homogeneous among the observations. This is a priori in contradiction with the findings of the global models using expansion terms. One could argue that the heterogeneity identified in the expansion models refers to the household heterogeneity, and not specifically to spatial heterogeneity, as it would have been had the attributes been expanded according to their coordinates (through the use of trend surface analysis for example).

In fact, some of the variables describing the household profile are not spatially structured, as indicate the Moran’s I values shown in Table 8. For the attributes that have been expanded with these “non-spatial” household characteristics, it is to be expected that they are not identified in the GWR framework as spatially heterogeneous (although other dimensions than household profile and preferences could be the cause of heterogeneity). However, both the income (Household Income and Income 80 K up) and the educational attainment of the households (University degree holders) do present a spatial structure, with significant Moran’s I values at the 95% confidence level. The attributes that are significantly expanded in the global models with these two characteristics should also be identified in the GWR models as heterogeneous, that is, with significant p values. This concerns the

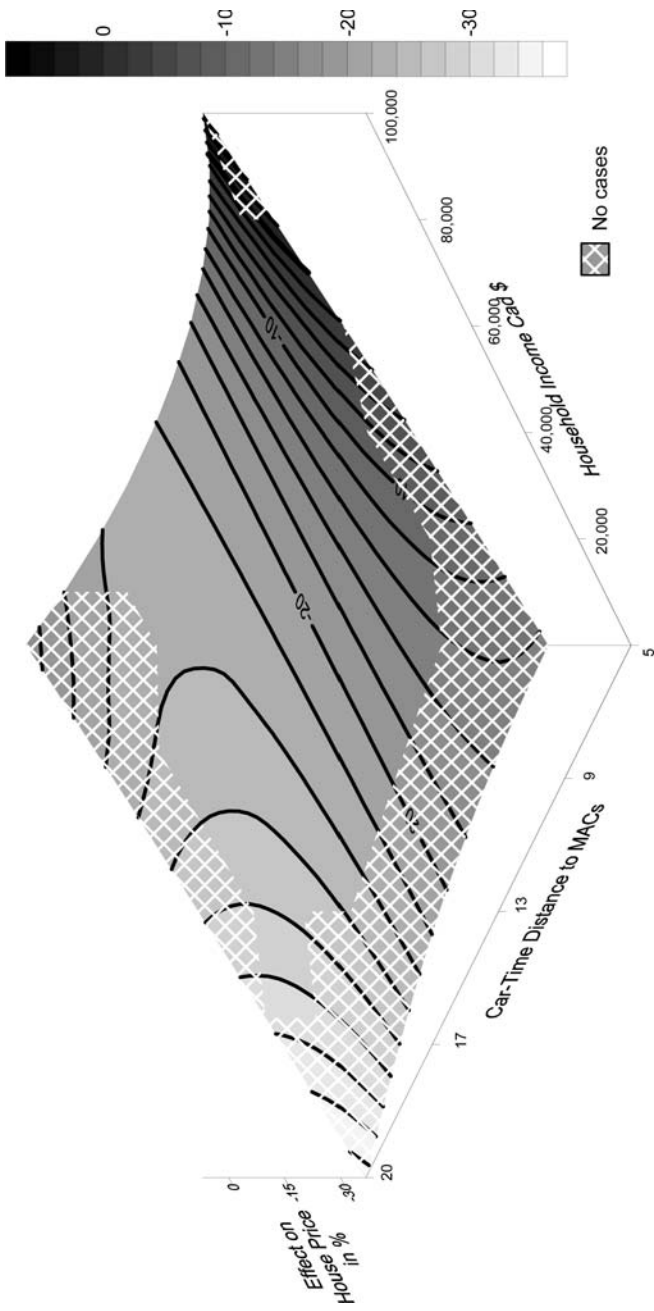


Fig. 5 Effect of car time distance to MACs considering household income

following: living area of a bungalow, in-ground pool, detached garage, car time to MACs and percentage of University degree holders. Whereas the two latter values are identified in the GWR as heterogeneous, the three former ones are not.

Concomitantly, two variables are considered heterogeneous within the GWRs, but are not significantly expanded in the global models (local tax rate) and NDVI Stdd. 1 km [land-use heterogeneity]). We can assume that the heterogeneity associated with these two attributes is not related to variations in the household profiles.

Both methods yield highly interesting results. Whereas spatial expansion makes it possible to consider both the spatial and the non-spatial heterogeneity of parameters, GWR provides interesting information through local regression statistics. However, although GWR is an interesting tool to identify and spatially describe non-stationary processes, it does not identify the cause of the parameter drift. Spatial expansion on the contrary, although less precise locally, makes it possible to investigate the cause of non-stationarity, thereby helping to disentangle the complex interactions influencing property values.

4.6 Some provocative findings

4.6.1 *Accessibility and income*

It is worthwhile to underline the significant drift of accessibility (car time to MACs, under its squared form) regarding household income. The car time to MACs is negatively linked to property values: each additional minute away from the city centre lowers the property value of 1.82%. However, this relation is not strictly linear but rather follows a U-shaped curve form, as shown by the significant integration of the squared form of the variable, with a positive sign. Furthermore, this squared term significantly interacts with the household income, with a positive sign too. This shows that the higher the income, the higher the squared term. Therefore, the devaluation associated with distance is more important for low-income than for high-income households, as shown on the three-dimensional surface of Fig. 5. This tends to corroborate the distance-cost trade-off theory, stating that high-income households can afford additional transportation costs and are ready to pay more for properties located in the outer-city limits. Also, the increasing practice of telework, which particularly concerns managers and professionals, may have an effect on the propensity of the most highly educated people to locate in more remote areas of the urban scene. In fact, those who can spend some working days at home may be willing to pay more for non-central locations, thus benefiting more from premium environments than typical commuters. These findings lead to further investigation using 2001 Origin-destination survey for Quebec City to analyze the spatial distribution of higher income telecommuters.

4.6.2 *Social homogeneity*

As foreseen, the percentage of university degree holders in the census tract has a global positive effect on the property value, each additional 10% adding a premium of 4.41%. This variable is among the most significant ones, with a t value of 9.17. Additionally, the expansion with the household-level binary variable “Holding a university degree” marginally proved significant, with an additional 1.81% premium. This shows that all things being equal, highly educated buyers who select single-family housing are willing to pay more in their quest for social homogeneity.

4.7 Summary and conclusions

This paper aims at understanding how the marginal value given to property and location attributes may vary among buyers. A telephone survey was conducted in order to obtain detailed information about 761 households that acquired single-family properties in Quebec, Canada, during the 1993–2001 period. Household-level attributes were introduced into hedonic functions to measure the effect of the homebuyer’s socio-economic context on implicit prices. Both the expansion method (Casetti 1972, 1997) and GWR (Fotheringham et al. 2002; Fotheringham et al. 1998) are used to assess the eventual heterogeneity of the impact of property specifics and location attributes.

A major finding is that some characteristics of the buyer’s household have a direct impact on property prices, namely the household income, the previous tenure status, and age. These findings must be put into the perspective of a specific location (Quebec City) and specific market conditions, that is, mainly a seller market with high supply and rather low demand for housing, at least for most of the period considered. Under these particular conditions, and using appropriate space-sensitive interaction methods, we could show that for each additional \$10,000 of income, a buyer pays a premium of 1.61% on average (+1.46 to 1.73%), all other things being equal. Also, the marginal effect of the household income is the fifth most significant parameter after the size (living area), the age of the property (apparent age), the social status of the neighbourhood (percentage of university degree holders in the Census tract), and accessibility (Car Time to MACs) (N3). Several hypotheses can explain the parameter significance and its positive sign. First, it is possible that the lack of descriptors defining the luxury attributes of the higher segment of the property market may result in a premium appearing as associated to the buyer’s income. However, as their ability to pay is increased, high-income buyers may also be less willing to engage in lengthy price negotiations, and may accept higher selling prices. Concomitantly, households with more restricted financial means may take more time to find the “best” deal as their budget is inflexible. While taking more time, they may visit more houses and thereby increase their chances to find sellers who on the contrary, have time constraints, and may want to sell rapidly. It would be interesting to obtain information about the seller’s profile, which can also be assumed to impact on the property sale price.

These findings should be compared with information on the time elapsed between the decision to look for a piece of property and the actual act of buying one. It is probable that potential buyers who are well off may be more prone to materialise their housing needs as budget constraints do not represent a serious impediment. Furthermore, the argument that the property price (as well as the desire to make an investment) is a criterion for buying is significantly more frequent on the part of low-income households (See Kestens et al. 2004).

First-time owners, that is, households that were previously tenants, “save” an average of 4.2% (3.88–4.18%, depending on the models) compared with former-owner households, all other things being equal. Again, first-time buyers may obtain a better price by waiting longer to close a deal, and former owners can afford a more substantial down-payment due to the sale proceeds from the previous home.

The age variable did enter in as such in one of the models (M3), however with a low t value. Furthermore, this criterion was dropped when additional expansion terms or Census data were included. Some collinearity may still be at stake here, and any direct interpretation about the direct link between age and price is therefore risky.

The integration of numerous expansion terms shows how the implicit prices of some property specifics and location attributes vary with the buyer’s household profile. These findings partly complete Starret’s statement (Starret 1981). He hypothesised that capitalisation of an attribute is only complete if the residents’ preferences are homogeneous. In fact, the significant drift of parameters according to the household characteristics shows that the capitalisation of an attribute does vary according to the household profile. Certain characteristics of the household profile are also significantly linked to the odds-ratio of mentioning certain property or neighbourhood choice criteria (See Kestens 2004), that is, to the household preference, as far as the choice criteria can be interpreted as a proxy for preference. Certain choice criteria are difficult to translate into measurable determinants of value. In fact, among those choice criteria for which the odds-ratio of being mentioned is linked to the household profile, few find their equivalent as expansion terms. For example, among the neighbourhood choice criteria, the odds-ratio of mentioning “Proximity to services” is significantly linked to age, household type, or income. Educational attainment has no impact on the propensity to mention this criterion. However, this paper suggests that the drift of the value assigned to accessibility to the MACs is linked to educational attainment, and not age, household type, or income. Similarly, this paper shows that the value given to vegetation in the close surroundings of the property varies significantly with age (Nb trees 29 up expanded by age 40 and over and NDVI 40 m expanded by Age 30–39). However, the odds-ratio to mention the presence of trees as a choice criterion is not linked to age but to the previous tenure status and the household type (for trees in the neighbourhood) and educational attainment and income (for trees on the property).

Although this paper has stressed that the marginal value of certain attributes varies with the household profile, the links between the coefficient’s drift and preference (or choice criteria) need further exploration.

Straightforward relations between stated choice criteria and heterogeneity of implicit prices could not be established.

More specifically, two significant expansion terms are worth underscoring. The first shows that the marginal value of accessibility varies with the household's income. Whereas the location rent is linearly negative for low-income households, it has more of the form of a U-shaped curve for high-income households, which tend to add a premium to remote locations, *ceteris paribus*. The recent and growing development of telework may be part of the explanation. In the U.S., home-based telework has grown nearly 40% since 2001, concerning some 23.5 million employees in 2003 (Pratt 2003). In Canada, the 2001 Census reported some 8% of teleworkers. Furthermore, a recent study showed that, out of a sample of salaried teleworkers working at home and using information technology like the internet, 60.6% hold a university degree (Tremblay 2003), this number being far above the national average [22.6% (Statistics Canada 2001)]. This paper's findings are coherent with the hypothesis that highly-educated teleworkers are prepared to pay a premium for remote locations, as compared with daily commuters. While additional research is needed on this issue, this paper's findings prove most relevant in a context of population aging, followed by widespread household relocation in well-serviced areas. While popular belief is to the effect that suburban house values are doomed to fall, the rapidly expanding trend for teleworking might slow down, and eventually reverse, this fatality.

Moreover, the insertion within the hedonic framework of origin-destination survey data, which provides detailed information on work, shopping and leisure trips, could further our understanding of this phenomena. In fact, the concomitant development of Information Technology and the trend toward more balanced relations between work and family redefines our notions and limits of space and location.

The interaction, too, between the effect of the percentage of university degree holders in the Census tract and the educational attainment of the buyer provides insight into social homogeneity processes. With a positive sign, this parameter indicates that highly-educated households do pay a premium to fulfill their quest for social homogeneity. This partially confirms Goodman and Thibodeau's (2003) hypothesis, that "Higher income households may be willing to pay more for housing (per unit of housing services) to maintain neighbourhood homogeneity" (p. 123). This paper showed it to be true regarding educational attainment, and not directly the household income, although these two dimensions are correlated.

Methodologically, the two methods that were used proved efficient. Expansion terms make it possible to analyse and to fully explain the cause of the parameter heterogeneity, whether its structure be spatial or not. Geographically Weighted Regressions provide additional insight by measuring local regression statistics. Some inconsistencies about non-stationary parameters were detected and need further investigation. Interesting developments of the GWR approach may in this sense prove useful, for example by considering spatial error autocorrelation in GWR models (Pàez 2002a, b). However, we feel that both methods are complementary rather than substitutes for each other, and that the use of additional methods such as seemingly unrelated regressions (SUR) (Knight et al. 1995; Zellner 1962)

may further our understanding of the complexity of property markets and urban dynamics.

References

- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Paper presented at 2nd symposium on information theory, Budapest
- Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht
- Anselin L (1995) Local indicators of spatial association—lisa. *Geogr Anal* 27:93–115
- Aten B, Heston A (2005) Regional output differences in international perspective. In: Kanbur R, Venables AJ (eds) Spatial inequality and development (Forthcoming). World Institute for Development Economics Research (WIDER), London
- Brunsdon C, Aitkin M, Fotheringham AS, Charlton M (1999) A comparison of random coefficient modeling and geographically weighted regression for spatially non-stationarity regression problems. *Geogr Environ Model* 3:47–62
- Brunsdon C, Fotheringham AS, Charlton M (2002) Geographically weighted summary statistics—a framework for localised exploratory data analysis. *Comput Environ Urban Syst* 26:501–524
- Brunsdon CF, Fotheringham AS, Charlton ME (1996) Geographically weighted regression: A method for exploring spatial non-stationarity. *Geogr Anal* 28:281–298
- Can A (1990) The measurement of neighbourhood dynamics in urban house prices. *Econ Geogr* 66:254–272
- Can A (1992) Specification and estimation of hedonic housing price models. *Reg Sci Urban Econ* 22:453–477
- Casetti E (1972) Generating models by the expansion method: Applications to geographical research. *Geogr Anal* 4:81–91
- Casetti E (1986) The dual expansion method: An application for evaluating the effects of population growth on development. *IEEE Trans Syst Man Cybern* 16:29–39
- Casetti E (1997) The expansion method, mathematical modeling, and spatial econometrics. *Int Reg Sci Rev* 20:9–32
- Cliff AD, Ord JK (1981) Spatial processes, models and applications. Pion Ltd., London
- Craven P, Wahba G (1979) Smoothing noisy data with spline functions. *Numerische Mathematik* 31:377–403
- Duda RO, Hart PE (1973) Pattern classification and scene analysis. Wiley and Sons, New York
- Fotheringham AS (2000) Context-dependent spatial analysis: A role for gis? *Geogr Syst* 2:71–76
- Fotheringham AS, Brundson FC, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, Chichester
- Fotheringham AS, Charlton ME, Brundson CF (1998) Geographical weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ Plan A* 30:1905–1927
- Geoghegan J, Wainger LA, Bockstael NE (1997) Spatial landscape indices in a hedonic framework: an ecological economics analysis using gis. *Ecol Econ* 23:251–264
- Getis A (1995) Spatial filtering in a regression framework: experiments on regional inequality, government expenditures, and urban crime. In: Anselin L, Florax RJGM (eds) New directions in spatial econometrics. Springer, Berlin Heidelberg New York, pp 172–188
- Getis A, Griffith DA (2002) Comparative spatial filtering in regression analysis. *Geogr Anal* 34:130–140
- Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. *Geogr Anal* 24:189–206
- Goodman AC, Thibodeau TG (1998) Housing market segmentation. *J Housing Econ* 7:121–143
- Goodman AC, Thibodeau TG (2003) Housing market segmentation and hedonic prediction accuracy. *J Housing Econ* 12:181–201

- Griffith DA (1988) *Advanced spatial statistics: special topics in the exploration of quantitative spatial data series*. Kluwer, Dordrecht
- Griffith DA (1996) Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *Can Geogr* 40:351–367
- Griffith DA (2000) A linear regression solution to the spatial autocorrelation problem. *J Geogr Syst* 2:141–156
- Hope ACA (1968) A simplified monte carlo significance test procedure. *J R Stat Soc Ser B* 30:582–598
- Hurvich CM, Simonoff JS, Tsai C-L (1998) Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *J R Stat Soc Ser B* 60:271–293
- Jaccard J, Turrisi R, Wan CK (1990) *Interaction effects in multiple regression*, vol 72. Sage Publications, Newbury Park
- Kelejian H (1995) Spatial autocorrelation: A suggested alternative to the autoregressive model. In: Anselin L, Florax RJGM (eds) *New directions in spatial econometrics*. Springer, Berlin Heidelberg New York
- Kestens Y (2004) Land use, accessibility and household profiles: Their effects on residential choice and house values. PhD Thesis, Université Laval
- Kestens Y, Thériault M, Des Rosiers F (2004) The impact of land use and surrounding vegetation on single-family property prices. *Environ Plan B Plan Des* 31:539–567
- Knight JR, Dombrow J, Sirmans CF (1995) A varying parameters approach to constructing house price indexes. *Real Estate Econ* 23:187–205
- Lancaster KJ (1966) A new approach to consumer theory. *J Pol Econ* 74:132–157
- Leung Y, Mei C-L, Zhang W-X (2000) Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environ Plan A* 32:9–32
- Loader C (1999) *Local regression and likelihood*. Springer, Berlin Heidelberg New York
- Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* 37:17–23
- Ord JK, Getis A (1995) Local spatial autocorrelation statistics. *Geogr Anal* 27:286–306
- Pàez A, Uchida T, Miyamoto K (2002a) A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. *Environ Plan A* 34:733–754
- Pàez A, Uchida T, Miyamoto K (2002b) A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests. *Environ Plan A* 34:883–904
- Pandit K, Casetti E (1989) The shifting patterns of sectoral labor allocation during development: Developed versus developing countries. *Ann Assoc Am Geogr* 79:329–344
- Payne BR (1973) The twenty-nine tree home improvement plan. *Nat Hist* 82:74–75
- Pratt J (2003) Teleworking trends: From sneakernet to internet. In: Paper presented at ITAC's annual conference, Baltimore, 4 september 2003
- Rosen S (1974) Hedonic prices and implicit markets: product differentiation in pure competition. *J Pol Econ* 82:34–55
- Schwartz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464
- Starret DA (1981) Land value capitalization in local public finance. *J Pol Econ* 89:306–328
- Statistics Canada (2001) Educational attainment, 2001 census, vol 2003. Statistics Canada
- Thériault M, Des Rosiers F, Joerin F, Villeneuve P, Lee-Gosselin M (2003) Modelling the impact of accessibility to services on housing prices: combining fuzzy logic and hedonic modelling within transportation gis. In: Conference tERES, Helsinki
- Thériault M, Des Rosiers F, Vandersmissen MH (1999) Gis-based simulation of accessibility to enhance hedonic modeling and property value appraisal, Joint IAAO-URISA 1999 conference, New Orleans, pp 151–160
- Tremblay D-G (2003) Telework: a new mode of gendered-segmentation? Results from a survey in canada. *Chaire Bell en Technologies et Organisation du Travail, Télé-Université, Québec*, pp 29
- Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Rem Sens Environ* 8:127–150
- Tueller PT (1989) Remote sensing technology for rangeland management applications. *J Range Manage* 42:442–452
- Tyrvaïnen L (1997) The amenity value of the urban forest: An application of the hedonic pricing method. *Landscape Urban Plan* 37:211–222

-
- Wu J, Gao W, Tueller PT (1997) Effects of changing spatial scale on the results of statistical analysis with landscape data: a case study. *Geogr Inform Sci* 3:30–41
- Zellner A (1962) An efficient model of estimating seemingly unrelated regressions and testing for aggregation bias. *J Am Stat Assoc* 57:348–368