Mass Appraisal, Hedonic Price Modelling and Urban Externalities: Understanding Property Value Shaping Processes

François Des Rosiers and Marius Thériault
Laval University, Canada

Advances in Mass Appraisal Methods Seminar,

Delft University of Technology,
October 30-31, 2006

Research funded by SSHRC

UNIVERSITÉ LAVAL

CRAD
Centre de Recherche en Aménagement et Développement

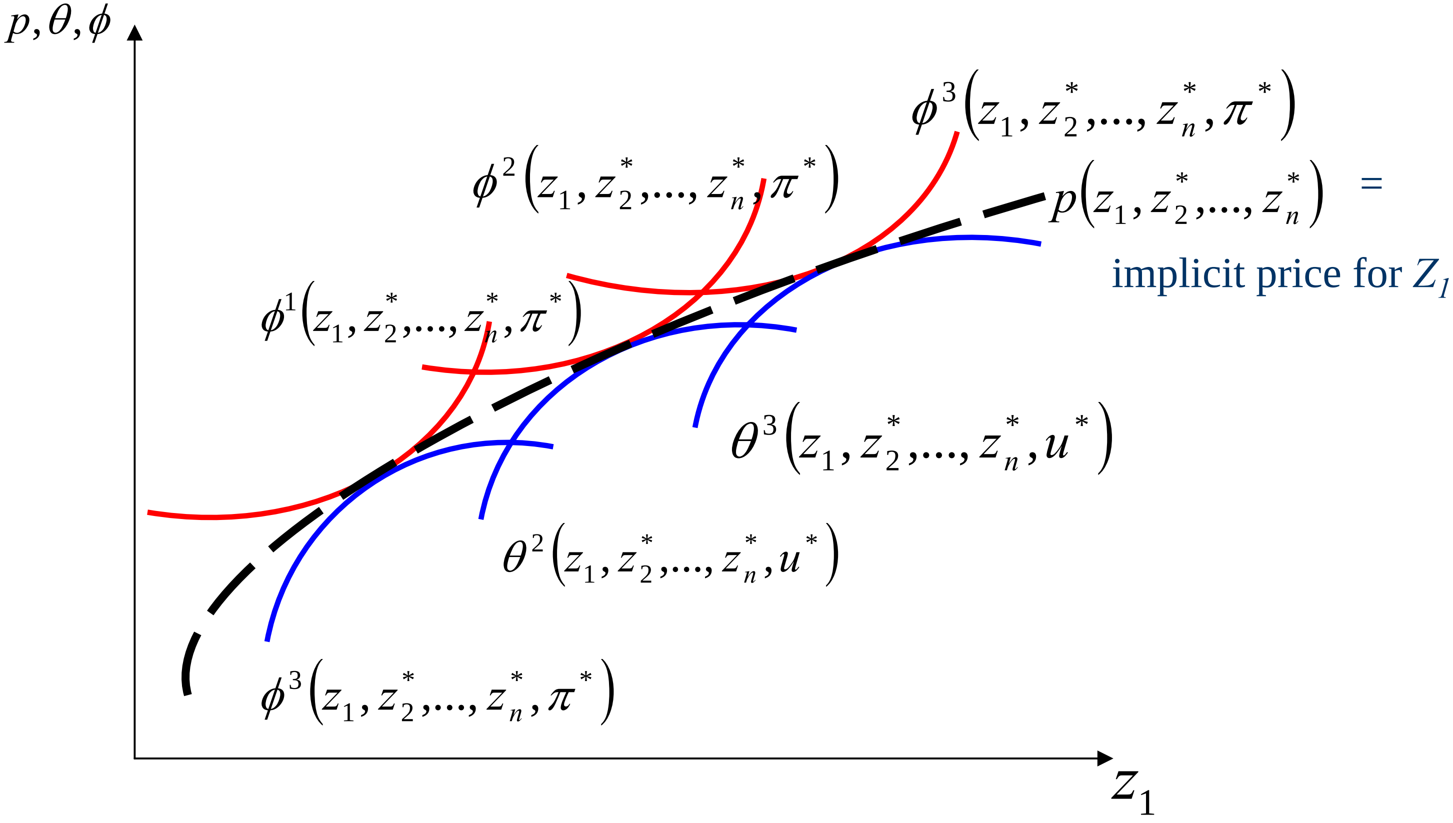
Why are hedonics so popular?

- It rests on multiple regression analysis (MRA), a conceptually sound and most powerful analytical device
- it perfectly fits the very definition of market value, expressed as "the most probable price" and, therefore, as a probability distribution
- The hedonic approach is not confined to producing value estimates as it adds most useful insights into the causal dimensions of property value determination

Why are hedonics so criticized (by heretics…)?

- It remains structurally bound to assuming *a priori* some functional relationship between sale prices and property attributes, based on either deductive or inductive grounds, or both
- Multicollinearity among variables as well as heteroskedasticity and spatial autocorrelation may result from not complying with restrictive conditions, thereby invalidating statistical tests
- Relationship between Y and X_i may not be linear
- Market analyzed may not be homogeneous over space

The conceptual framework



Measuring proximity effects - HVTL

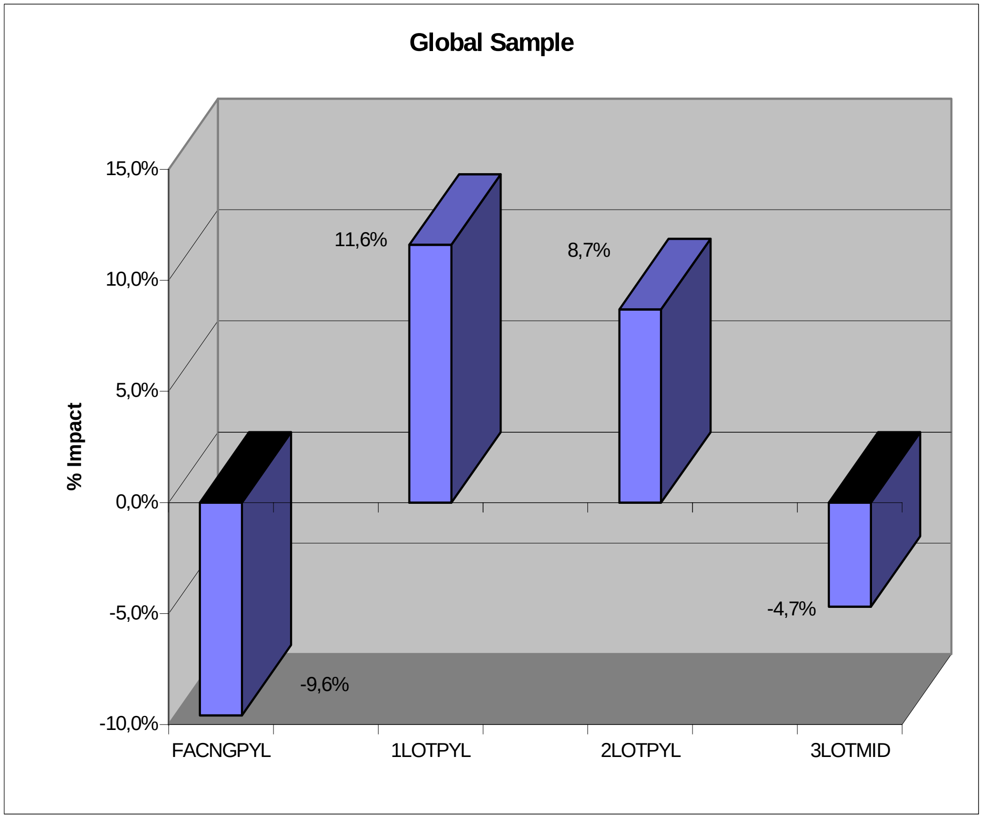
- 507 single-family houses sold over the 1991-96 period in the City of Brossard (pop.: 69,000 by 1996), located in the Greater Montreal area, Canada
- The study area is between 250 and 500 metres wide and is bounded by three major highways, with a 315 Kv. transmission line running through its centre.
- Asymmetrical location of the line: within 50 metres of the eastern boundary of the easement, as opposed to 15 metres on the west side
- Overall, 383 houses have a limited, moderate or pronounced rear, side or front view on the line, with 34 being directly adjacent to it.

Measuring proximity effects - HVTL

- 25 property descriptors pertaining to physical, neighbourhood, environmental, access, fiscal and sales time attributes
- A series of HVTL-related descriptors: linear distance to the line and easement as well as dummy distance variables (50 and 100 m. increments)
- Dummy variables to control for pylons' position relative to houses that are adjacent to the easement (house facing pylon, located one, two or three lots away from pylon, or mid-span located)

Main findings

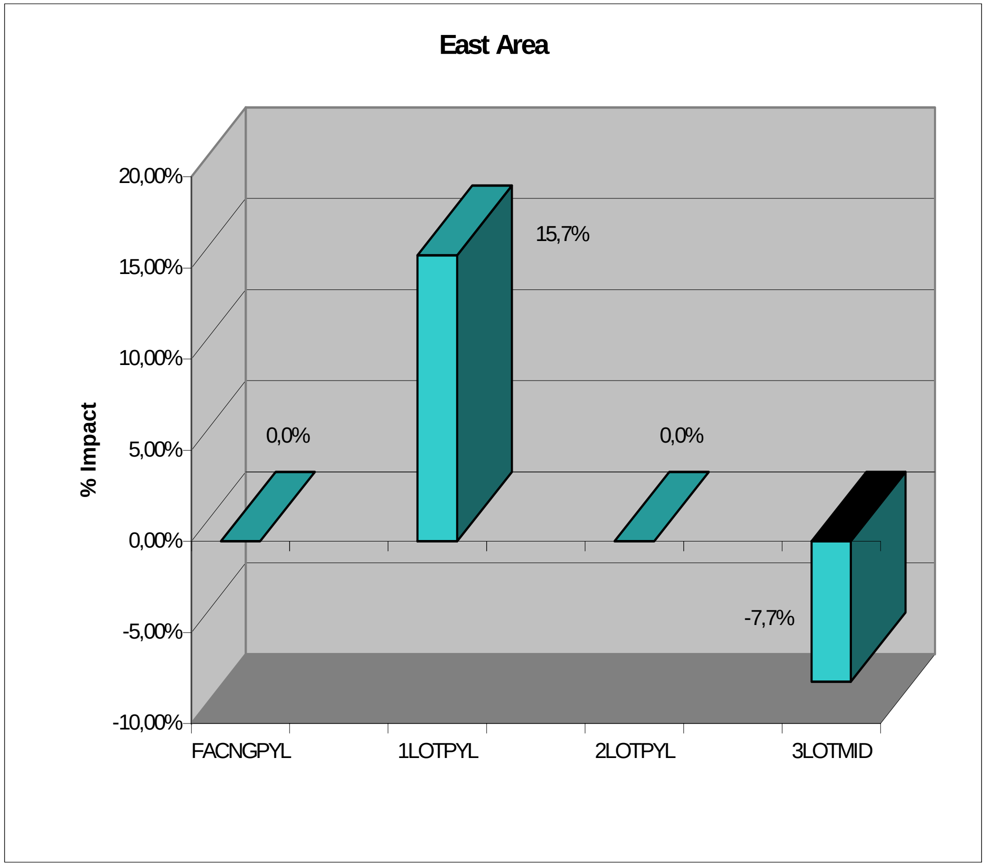
	HVTL Attribute	% Impact
Global Sample		
House facing pylon:	FACNGPYL	-9,6%
One lot away from pylon:	1LOTPYL	11,6%
Two lots away from pylon:	2LOTPYL	8,7%
Three lots away from pylon or mid-span location:	3LOTMID	-4,7%



Main findings

East Area (150 ft. setback to HVTL)

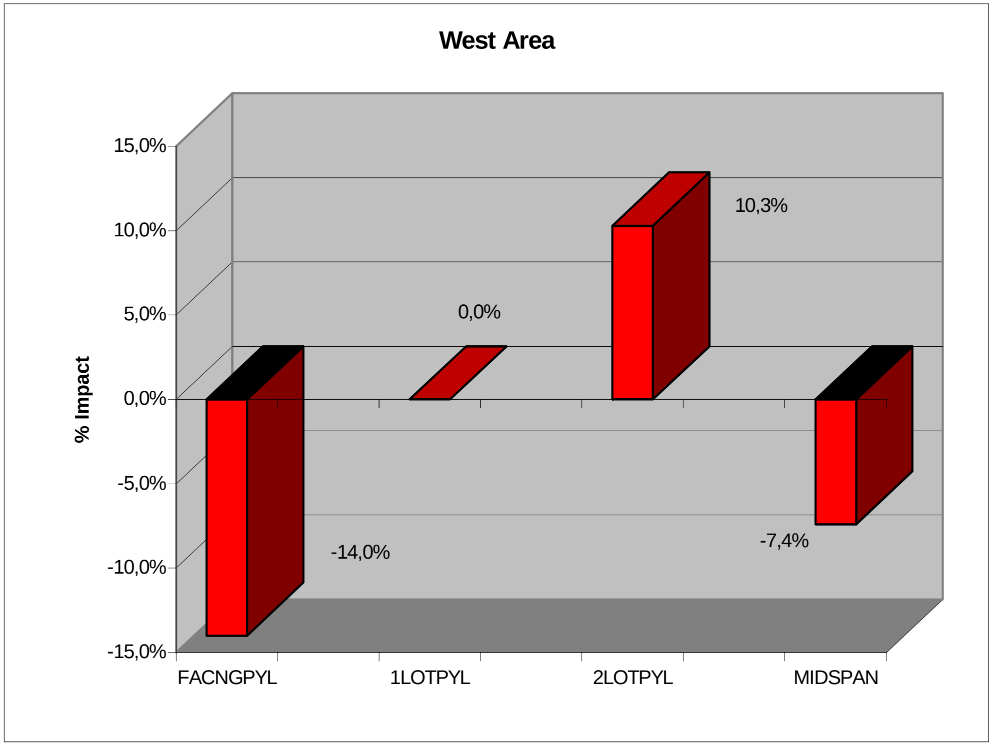
House facing pylon:	FACNGPYL	n.s.
One lot away from pylon:	1LOTPYL	15,7%
Two lots away from pylon:	2LOTPYL	n.s.
Three lots away from pylon or mid-span location:	3LOTMID	-7,7%



Main findings

West Area (50 ft. setback to HVTL)

House facing pylon:	FACNGPYL	-14,0%
One lot away from pylon:	1LOTPYL	n.s.
Two lots away from pylon:	2LOTPYL	10,3%
Mid-span location:	MIDSPAN (sig. 0.07)	-7,4%



N.B.:
Percentage price impacts reported here are an average of all significant coefficients derived from various functional forms and should therefore be viewed as indicators only. Besides, they reflect "gross" location impacts due to a view on pylons and conductors alone.

Handling non-monotonicity proximity to primary schools

- Easy access to a nearby school remains an overwhelming advantage for households with school-age children
- Too great proximity may be drive house prices down because of traffic, noise and, eventually, risk of damage to property
- An optimal distance from school should then exist, whereby the net positive impact on house value is maximized
- A similar rationale could be applied to school size, with both small and large school displaying advantages over middle-size institutions

Handling non-monotonicity proximity to primary schools

- Database: 4,300 bungalows (one-story, single-family detached houses) sold on the territory of the Quebec Urban Community (QUC, pop.: 675,000 by the time of study) between January 1990 and December 1991
- 116 elementary schools - three size categories: small (200 pupils and below), medium (201-500 pupils) and large ones (over 500 pupils)

Handling non-monotonicity proximity to primary schools

- The gamma distribution is a probability density function given by:

$$\mathbf{f\,(x) \quad = K * x^{(\alpha-1)}\, e^{(-x\,/\,\beta)}} \quad \text{for } x > 0$$

$$\mathbf{= 0} \quad \text{for } x = 0,$$

where α and β are positive parameters and K is a constant.

- For specific values of the parameters α and β, the gamma distribution turns into an exponential distribution, a chi-square distribution or even approaches a normal distribution.
- Moreover, the lower the value of β the steeper the slope beyond the maximum

Handling non-monotonicity proximity to primary schools

- **LnSALEPRICE = Ln K1 + (α_1-1) Ln DSCHOOL**
  - DSCHOOL / β_1
  + Ln Φ (SCHLSIZE) + Σ Bi Zi + e
- The first derivative of the gamma function set to zero provides a measure of the "optimal" distance away from a nearby school in order for a property to have its value maximized. Thus, we can write:

  ***d* LnSALEPRICE / *d* DSCHOOL**

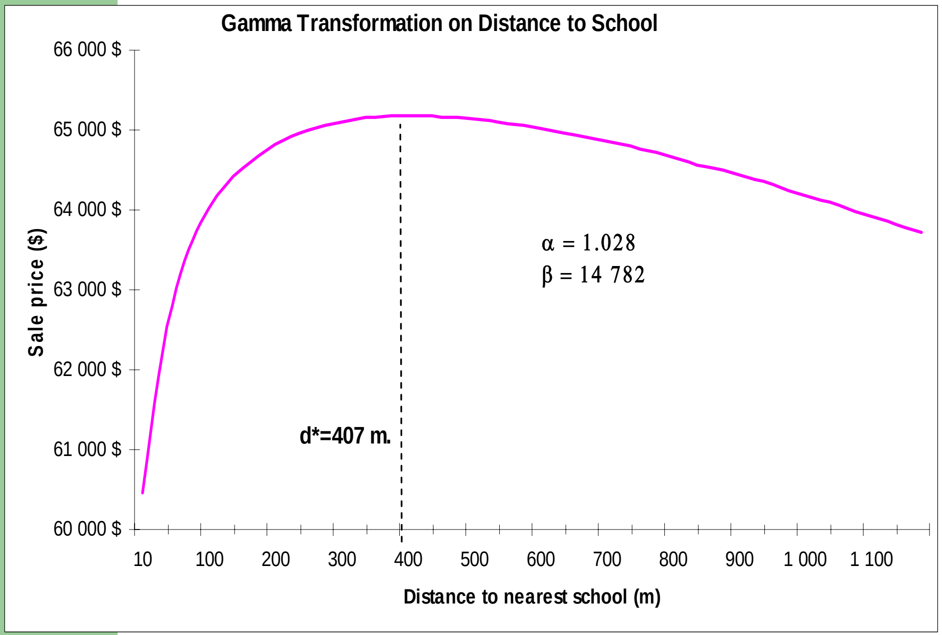
  = (α_1-1) * 1/ DSCHOOL - 1 / β_1
- Hence: **DSCHOOL* = (α_1-1) β_1**

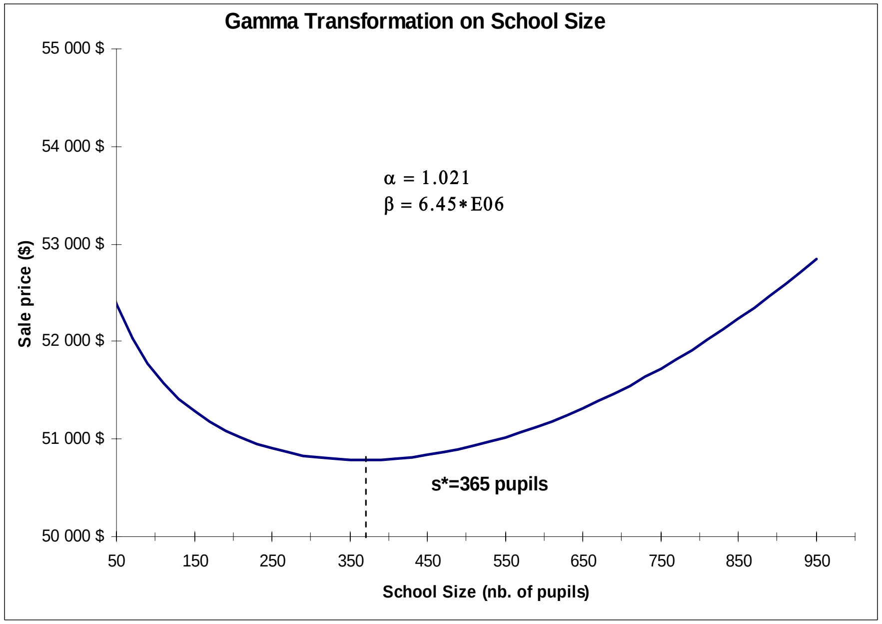
Handling non-monotonicity proximity to primary schools

- Due to excessive collinearity, a modified gamma function is applied to the price-size relationship
- We end up with a double-gamma transformation on distance and size

Name of Variable	Regression Coefficient	T value	Probability > |T|
LnDSCHOOL* **	0.0275	5.62	0.000
DSCHOOL* **	-0.0001	-11.08	0.000
LnSCHLSIZE **	-0.0207	-2.17	0.030
SCHLSIZE2 **	0.0000	2.09	0.036
	Alpha	**Beta**	
Gamma parameters on Distance-to-school	**1.03**	**14 782**	***Optimal Distance (m.)*** ***407***
Gamma parameters on School Size	**1.02**	**6.45E+06**	***Value Minimizing Size (# pupils)*** ***365***

Handling non-monotonicity proximity to primary schools





Modelling Accessibility to Urban Services

- The accessibility potential of any location – or attraction point - is usually expressed as a direct function of the number of opportunities it offers as a destination for households while being inversely related to its distance (or travel time) to residential places
- However, this “objective”, supply-driven definition is increasingly challenged by researchers
- “Subjective”, demand-driven accessibility may vary according to the type of amenity considered (*e.g.* workplaces versus leisure places) as well as among social groups

PCA-derived, supply- driven accessibility to urban services

- For each property in the database, best route (shortest trip duration) is computed to main employment centres, to schools, colleges and university as well as to neighbourhood, local and regional shopping centres.
- The computation algorithm identifies 52,500 street segments (acting as directional links) and 19,250 nodes (acting as street intersections)
- Distances and travel times by car and on foot to the nearest amenity are computed for every street node

PCA-derived, supply- driven accessibility to urban services

- 15 accessibility attributes are defined, and then grouped into factors using PCA, with a Varimax rotation
- 2 factors are retained that explain 75% of the variations in the data:
- F1 = access to regional services
- F2 = access to local services

Total Variance Explained

	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.683	64.556	64.556	6.313	42.086	42.086
2	1.668	11.122	75.678	5.039	33.592	75.678

Extraction Method: Principal Component Analysis.

Estimating demand-driven accessibility using fuzzy logic

- An O-D survey is used to identify suitability thresholds for daily commuters; it is assumed that:
- [1] any travel time smaller than the observed median (*C50*) during the O-D survey leads to an acceptable destination
- [2] a travel time larger than the 90th percentile of actually reported trip (*C90*) is likely to be unsatisfactory
- [3] intermediate cases yield satisfactory levels obtained through linear interpolation (fuzzy membership).

Estimating demand-driven accessibility using fuzzy logic

- Computing the sum of suitability indices over every service locations assesses the raw suitability of each residential location
- Values are then rescaled between 0 and 100, using the city-wide maximum local raw suitability value as the denominator
- This procedure is applied to every set of trip purpose and type of individual or household reported in the O-D survey

Controlling for urban centrality

$$\mu_{ij} = \frac{e^{\lambda} P_i^{\lambda_o^p} P_j^{\lambda_d^p}}{D_{ij}^{\lambda^d}}$$

μ_{ij} = Expected number of car trips between locations *i* and *j*

P_i = Total population at residential location *I*

P_j = Total number of potential activities at location *j*

Dij = Travel time by car from residential location *i* to activity location *j* (minutes)

Modelling accessibility

	Model 1				Model 2				Model 3			
	Unstdz. B	**Std. Error**	**Standzd. Beta**	***t* Value**	**Unstdz. B**	**Std. Error**	**Standzd. Beta**	***t* Value**	**Unstdz. B**	**Std. Error**	**Standzd. Beta**	***t* Value**
(Constant)	11.68731	.04746		246.3	11.55619	.05250		220.1	11.50028	.05038		228.3
LotSize (m^2)	.00003	.00002	.031	1.6	.00008	.00002	.078	4.3	.00008	.00002	.080	4.2
Bungalow * Living Area	.00235	.00018	.357	13.3	.00231	.00016	.351	14.4	.00228	.00017	.346	13.6
Cottage * Living Area	.00249	.00013	.569	19.4	.00250	.00012	.571	21.3	.00247	.00012	.565	20.2
Attached * Living Area	.00149	.00027	.101	5.5	.00098	.00025	.067	3.9	.00112	.00026	.076	4.3
Apparent Age	-.00387	.00057	-.138	-6.8	-.00853	.00062	-.303	-13.8	-.00662	.00060	-.235	-11.0
# Washrooms	.09517	.01252	.144	7.6	.07281	.01151	.110	6.3	.08150	.01197	.124	6.8
#Fireplace	.05082	.01209	.079	4.2	.04970	.01101	.077	4.5	.05106	.01149	.079	4.4
Hard Wood Stair	.07454	.01623	.096	4.6	.05922	.01489	.076	4.0	.06646	.01544	.085	4.3
High Quality Floor	.06689	.01295	.097	5.2	.04912	.01185	.071	4.1	.05814	.01232	.084	4.7
LargeTerrace	.12394	.04813	.045	2.6	.10813	.04382	.039	2.5	.10856	.04577	.039	2.4
Brick Ext. Walls (≥51%)	.04567	.01420	.064	3.2	.03660	.01294	.051	2.8	.04089	.01349	.057	3.0
Clapbord Ext. Walls (≥51%)	-.05414	.01565	-.069	-3.5	-.04675	.01425	-.060	-3.3	-.05210	.01489	-.067	-3.5
Single Attached Garage	.13307	.02731	.085	4.9	.11599	.02488	.074	4.7	.12187	.02598	.078	4.7
Double Attached Garage	.16945	.03793	.080	4.5	.13446	.03459	.063	3.9	.15802	.03626	.074	4.4
Double Detached Garage	.10959	.03132	.062	3.5	.12144	.02857	.069	4.3	.11030	.02974	.062	3.7
Excavated Pool	.18383	.02617	.125	7.0	.16487	.02386	.112	6.9	.16491	.02495	.112	6.6
Month93Jan	-.00184	.00045	-.070	-4.0	-.00167	.00041	-.063	-4.1	-.00191	.00043	-.072	-4.5
OvTaxRate	-.25656	.01589	-.292	-16.1	-.14557	.02068	-.166	-7.0	-.25032	.01575	-.285	-15.9
Acces_Factor1 (Reg. services)					***.12485***	***.00959***	***.322***	***13.0***				
Acces_Factor2 (Local services)					***.04177***	***.00871***	***.090***	***4.8***				
AWork * NoWorkerHld									***.00287***	***.00042***	***.181***	***6.8***
AWork * WorkerHld									***.00273***	***.00035***	***.216***	***7.7***
Centrality Index									***.00173***	***.00051***	***.068***	***3.4***

Model	Accessibility / Centrality Index	R Square	SEE	Unstdz. B	Standzd. Beta	*t* Value	VIF
3 **Workplaces * Hsld Profile**	AWork * NoWorkerHld	.758	.1704	.00287	.181	6.8	2.752
	AWork * WorkerHld			.00273	.216	7.7	3.061
	Centrality Index			*.00173*	*.068*	*3.4*	*1.575*
4 **Schools * Family Status**	ASchool * Family	.765	.1678	.00333	.279	9.6	3.431
	ASchool * ChildlessHld			.00255	.220	8.0	3.068
	Centrality Index			*.00146*	*.058*	*2.9*	*1.572*
5 **Large Shops * Family Status**	ALargeShop * Family	.758	.1702	.00230	.186	7.9	2.200
	ALargeShop * ChildlessHld			.00235	.138	6.0	2.059
	Centrality Index			*.00172*	*.068*	*3.4*	*1.581*
6 **Small Shops**	ASmallShop	.759	.1698	.00276	.168	8.2	1.655
	Centrality Index			*.00152*	*.060*	*3.0*	*1.616*
7 **Groceries * Family Status**	AGrocery * Family	.756	.1710	.00257	.185	7.3	2.479
	AGrocery * ChildlessHld			.00222	.130	5.4	2.281
	Centrality Index			*.00157*	*.062*	*3.0*	*1.685*
8 **Leisure * Family Status**	ALeisure * Family	.762	.1689	.00290	.242	8.8	3.001
	ALeisure * ChildlessHld			.00272	.193	7.3	2.804
	Centrality Index			*.00143*	*.056*	*2.8*	*1.618*
9 **Health care * Family Status**	AHealthCare * Family	.766	.1673	.00342	.265	9.9	2.947
	AHealthCare * ChildlessHld			.00262	.199	7.9	2.574
	Centrality Index			*.00124*	*.049*	*2.4*	*1.618*
10 **Restaurants**	ARestaurant	.768	.1668	.00323	.212	10.1	1.801
	Centrality Index			*.00120*	*.047*	*2.4*	*1.608*
11 **Workplaces * Age Groups**	AWork * Age34less	.757	.1704	.00220	.155	5.9	2.698
	AWork * Age35-44			.00301	.306	9.0	4.507
	AWork * Age45-54			.00324	.318	9.7	4.236
	AWork * Age55more			.00317	.194	8.1	2.229
12 **Workplaces * Hsld Income**	AWork	.771	.1655	.00311	.179	8.3	1.914
	AWork * Income<60K$			-.00111	-.098	-4.7	1.811
	AWork * Income60-80K$			-.00060	-.050	-2.5	1.682
	AWork * Income80-100K$			-.00029	-.021	-1.1	1.544
	AWork * income>100K$			.00074	.060	2.9	1.737
	Centrality Index			*.00192*	*.076*	*3.9*	*1.582*

Dealing with spatial dependence

- In traditional hedonic price modelling, the contextual variations over space are usually specified using “fixed” coefficients – derived from location dummy variables - to assess their direct effect on house values
- This is based on the assumption that the marginal prices of structural housing attributes are invariant through space
- Market heterogeneity is a major source of spatial autocorrelation among residuals if not adequately handled in the model

Dealing with spatial dependence

- Spatial autocorrelation may be defined as an average correlation between observations based upon replicated realisations of the geographic distribution of some attribute (Griffith 1992)
- Exogenous effects can actually be manifold, ranging from city-wide structural factors to local externalities
- Two approaches are used here to deal with spatial dependence:

  [1] Casetti's Spatial Expansion Method (SEM)

  [2] Geographically Weighted Regression (GWR)

Spatial expansion method (SEM)

- Essentially, the SEM “extends” fixed parameters by introducing interactive variables combining a previously defined fixed characteristic with a context-sensitive, space-dependent variable.
- The hedonic equation may then be expressed as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + ((\mathbf{CE})\mathbf{X}^{t})\mathbf{1} + \varepsilon$$

where the second, expanded term accounts for interactions between basic housing attributes and context-sensitive variables (neighbourhood or household-related features).

Geographically weighted regression (GWR)

- With the GWR approach, moving regression functions are estimated for every sampling point in a regular grid, using all data within a certain region around this point for calibration
- The resulting parameters are site-specific and can therefore vary through space
- A weighting scheme may be designed, whereby a spatial kernel is applied in order to give greater influence to close data points
- The spatial kernel may be fixed (identical for all locations) or adaptive, in which case its bandwidth will vary with the density of the data

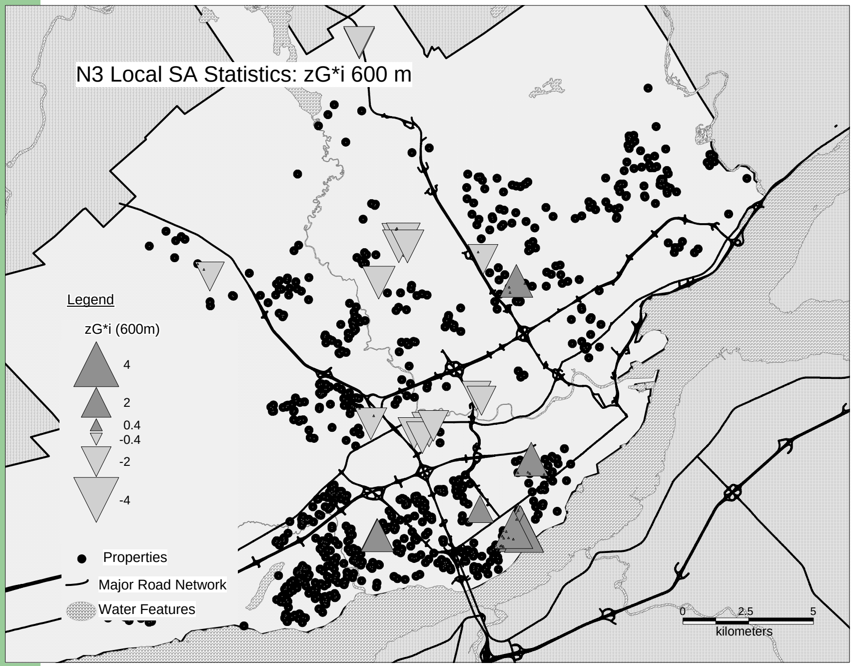
Comparing the two approaches

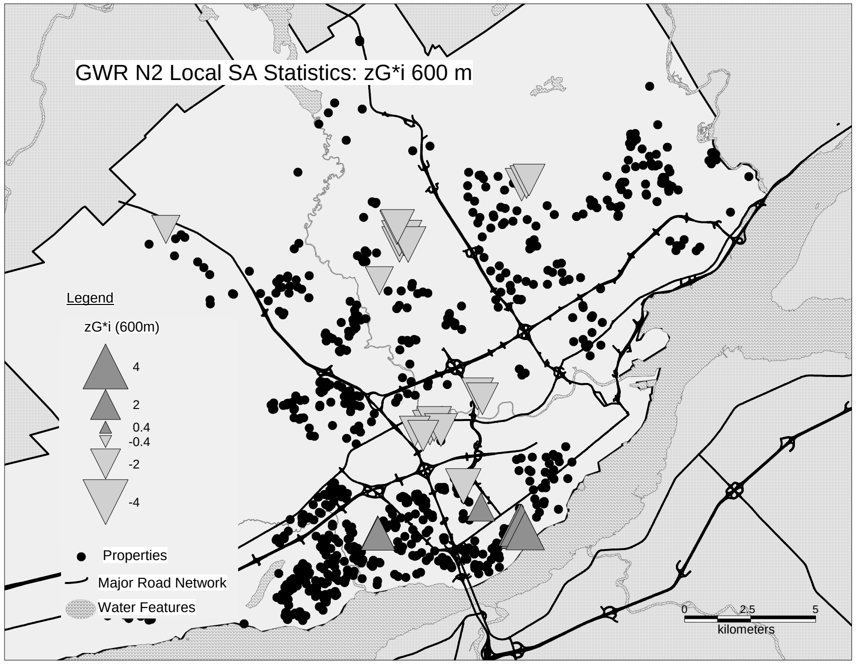
- SEM and GWR are applied to a sample of 761 single-family houses sold between 1993 and 2001 (between 1993 and 1996 mainly) in Quebec City, Canada (Kestens *et al.*, 2006)
- In addition to basic land, building and local tax features, models control for several other dimensions, namely:
- [1] accessibility to urban services, expressed as mean time-distance by car to main activity centres (MAC)
- [2] surrounding vegetation
- [3] information on buyer's household profile obtained through a phone survey carried out from 2000 to 2003
- [4] socio-economic and housing stock information derived from Census data.

Comparing the two approaches

Dependent Variable: Ln Sale Price Number of cases = 761		
Variables in Model	Property specifics Accessibility Land use and Vegetation in buffers around each property Buyer's Household-level attributes 1996 Census data (Enumeration area-level) For OLS Model: Interactions (Household attributes * others variables)	
OLS / SEM Model		
Model Adjustment	R-square	0.894
	Adj. R-Square	0.889
	SEE	0.104
	SEE in %	10.9%
	F ratio	161
	Sig.	0.000
	Df1/Df2	38/722
	Interactive Variables / Total Variables	11/38
	Maximum Variance Inflation Factor value	3.9
Spatial Auto-correlation of Residuals	Moran's I (within 1500 m lag)	0.102
	Sig.	0.218
	Most sig. Moran's I SA range (300 m lags)	600-900
	Nb of significant LISA zG*i statistics (600 m lag, sig. 0.05)	26
	Nb of significant LISA zGi statistics (600 m lag, sig. 0.05)	17
GWR Hedonic Model		
Model Adjustment	R-square	0.892
	SEE	0.1059
	Kernel bandwidth (meters)	706.5
	F statistic of GWR Improvement (sig.)	2.51 (0.013)
Spatial Auto-correlation of Residuals	Moran's I (within 1500 m lag)	0.082
	Sig.	0.265
	Nb of significant LISA zG*i statistics (600 m lag, sig. 0.05)	26
	Nb of significant LISA zGi statistics (600 m lag, sig. 0.05)	20

Significant zG*i statistics for SEM & GWR Hedonic Model





Comparing the two approaches

- Both methods yield highly interesting results and leads to the conclusion that social and spatial heterogeneity, while linked to one another, are not strictly equivalent
- SEM makes it possible to consider both the spatial and the non-spatial heterogeneity of regression parameters
- GWR provides interesting information through local regression statistics but does not allow identifying the process behind the parameter drift
- Casetti's SEM, on the contrary, while less precise locally, makes it possible to explicitly consider actual processes lying at the root of non-stationarity

Concluding Comments

- MRA-based hedonic approach is a most powerful, highly versatile and adaptable method
- It requires market intelligence