



Canadian Journal of Regional Science
Revue canadienne des sciences régionales

Le choix de la forme fonctionnelle dans la théorie hédonique : retour sur un vieux débat

Jean Dubé, François Des Rosiers, et Marius Thériault

Respectivement : Département sociétés, territoires et développement, Université de Québec à Rimouski (UQAR) ; Faculté des sciences de l'administration, Université Laval ; École supérieure d'aménagement et de développement, Université Laval. Adressez vos commentaires à jean_dube@uqar.qc.ca.

Soumis 9 Septembre 2010. Accepté 17 Mai 2011.

© Canadian Regional Science Association/ Association canadienne des sciences régionales 2011.

Dubé, J, Des Rosiers, F, & Thériault, M. 2011. Le choix de la forme fonctionnelle dans la théorie hédonique : retour sur un vieux débat. *Canadian Journal of Regional Science / Revue canadienne des sciences régionales* 34(2-3), 81-90.

Cet article reprend un vieux débat lié au choix de la forme fonctionnelle paramétrique adaptée pour spécifier l'équation de prix hédonique. Bien que la transformation proposée par Box & Cox (1964) soit générique et permette de normaliser les distributions des variables dépendante et indépendantes, son utilisation a été très limitée durant les dernières décennies, cédant la place aux formes fonctionnelles classiques de type log-linéaire, semi-log et linéaire. Grâce à une série de tests portant sur le pouvoir prédictif de plusieurs formes de modèle et sur le comportement de leurs résidus, cette recherche vise à vérifier s'il est possible d'identifier, avec les marchés résidentiels unifamiliaux de la ville de Québec comme étude de cas, une forme fonctionnelle supérieure aux autres. Les résultats montrent que la spécification log-linéaire demeure, parmi les alternatives étudiées, une forme fonctionnelle paramétrique intéressante étant donné ses performances et sa simplicité d'interprétation. Par ailleurs, des régressions auxiliaires sur les résidus permettent d'identifier quelques problèmes liés à l'estimation des coefficients avec l'approche des moindres carrés ordinaires. L'article se termine par une réflexion sur la méthode d'estimation qui devrait être explorée dans les recherches futures.

In this paper, an old debate on the choice of the appropriate parametric functional form for modelling the hedonic price equation is revisited. The generic transformation proposed by Box & Cox (1964) provides a convenient way to normalize distributions of dependent as well as independent variables. However, it has been seldom used in the last few decades, with classic functional forms blossoming : log-linear, semi-log and linear. Using the Quebec City single-family residential property market as case study, this paper compares the several functional forms using a set of tests on predictive power as well as on structure of their residuals. Findings show that the log-linear form remains a quite interesting parametric functional form when compared with tested alternatives, considering its performance and the easy interpretation of its coefficients. However, using auxiliary regressions on residuals, we still detect problems related to the estimation of the coefficients with the ordinary least square method. We conclude by identifying alternative estimation methods that should be explored in future research.

La théorie hédonique a été formellement définie par Rosen (1974) dans un article phare qui a influencé la majorité des études empiriques subséquentes. Elle permet d'exprimer le prix de vente d'un bien complexe (automobile, propriété foncière, etc.) en fonction de chacun de ses attributs, tant physiques que environnementaux. Par contre, la théorie ne fournit aucune indication, a priori, sur la forme que

devrait prendre la fonction de prix hédonique. Si l'identification de la forme de la relation reliant le prix de vente aux attributs d'une propriété demeure une question fondamentale, elle est néanmoins négligée dans plusieurs applications empiriques.

L'approximation paramétrique optimale pour déterminer les prix hédoniques demeure peu documentée,

tant dans la littérature que pour le territoire de la ville de Québec. Au plan empirique, les formes fonctionnelles log-linéaire¹ et semi-log² sont très courantes, alors que la spécification linéaire est de moins en moins utilisée. Toutefois, le choix de la forme de la relation entre le prix de vente et les attributs du bien influence les résultats d'estimation. Cet article propose un retour sur ce vieux débat qui a été négligé depuis quelques décennies. À partir d'une étude de cas pour les marchés résidentiels unifamiliaux de la ville de Québec, l'objectif de cet article consiste à comparer diverses formes fonctionnelles paramétriques afin d'identifier celle (ou celles) qui est la plus pertinente pour réaliser les modélisations hédoniques. Bien que les résultats obtenus pour la ville de Québec soient difficilement généralisables à d'autres territoires, la méthodologie présentée est utilisable pour identifier la (ou les) forme(s) fonctionnelle(s) applicable(s) à d'autres marchés ou à d'autres territoires.

La méthode consiste à estimer les formes fonctionnelles paramétriques les plus utilisées dans la littérature, sous une spécification simple ou quadratique, en plus de la formulation la plus générale, fondée sur la transformation de Box-Cox. Ensuite, une série de tests portant sur le pouvoir prévisionnel des modèles et sur le comportement des résidus permettent de comparer les performances respectives des diverses formes fonctionnelles. Les résultats suggèrent des pistes de réflexion pour des recherches futures afin de développer une approche de modélisation paramétrique optimale, évitant ainsi de recourir aux formes semi-paramétriques ou non-paramétriques.

Le reste de l'article est divisé en six sections. La première présente brièvement quelques études portant sur le problème du choix de la forme fonctionnelle paramétrique, tout en reprenant les conclusions dégagées par les auteurs. La seconde présente la méthodologie retenue pour estimer les diverses formes fonctionnelles, en plus d'expliquer les tests effectués afin de mesurer les performances prévisionnelles du modèle et d'étudier la

Encadré 1.

$$P_{it}^{\lambda} = \alpha_0 + \sum_{k=1}^K \alpha_k Z_{kit}^{\theta} + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \gamma_{kl} Z_{kit}^{\theta} Z_{lit}^{\theta} + \varepsilon_{it} \quad (1)$$

$$P_{it}^{\lambda} = (P_{it}^{\lambda} - 1) / \lambda, \quad \lambda \neq 0, \quad (2)$$

$$= \ln P_{it}, \quad \lambda = 0$$

$$Z_{kit}^{\theta} = (Z_{kit}^{\theta} - 1) / \theta, \quad \theta \neq 0, \quad (3)$$

$$= \ln Z_{kit}, \quad \theta = 0$$

$$\hat{\varepsilon}_{it} = \pi_0 + \pi_1 P_{it}^{\lambda} + u_{it} \quad (4)$$

structure des résidus. La section suivante présente les données utilisées pour estimer les modèles empiriques du cas d'étude, avant d'analyser les résultats issus des spécifications identifiées. Ensuite, une brève discussion suggère une forme fonctionnelle appropriée pour le marché étudié, tout en révélant les lacunes de cette dernière, en plus d'identifier des pistes de recherche complémentaire. Finalement, une brève conclusion termine l'article.

Revue de littérature

La théorie hédonique, formellement présentée par Rosen (1974), permet d'exprimer le prix de vente d'un bien complexe en fonction de chacun de ses attributs, tant physiques qu'environnementaux. Grâce à la mise en œuvre de techniques de régression, elle permet de déterminer le prix implicite (ou hédonique) de chacune des caractéristiques du bien considéré. Bien que plusieurs applications empiriques proposent une équation de prix linéaire dans les paramètres, Rosen mentionne, dans son article, l'absence d'a priori sur la forme de la relation. Il souligne d'ailleurs que celle-ci n'est pas forcément linéaire. À son tour, Can (1992) souligne la complexité de l'identification de la forme fonctionnelle idéale dans la théorie hédonique, notamment en raison du manque de base conceptuelle.

Goodman & Thibodeau (1995) suggèrent l'utilisation d'une spécification non-linéaire de l'équation de prix

hédonique afin de permettre aux prix des caractéristiques de varier en fonction des autres attributs du bien complexe. Il existe plusieurs manières de généraliser des effets non-linéaires dans les variables d'un modèle de régression linéaire multiple. Des propositions relativement simples mobilisent des formes fonctionnelles qui utilisent des variables croisées et des transformations quadratiques.³

Une forme générale de la régression linéaire repose sur la transformation mathématique non-linéaire de variables continues. L'approche de Box & Cox (1964) généralise plusieurs formes fonctionnelles, ce qui permet de comparer les propriétés de diverses spécifications et de déterminer leurs pertinences respectives. Goodman (1978) est parmi les premiers à utiliser cette approche pour une application portant sur les transactions immobilières. Blackley *et al* (1984) suggèrent d'utiliser la transformation de Box-Cox afin d'évaluer l'ensemble des formes fonctionnelles possibles et d'identifier la meilleure option. Ils argumentent que cette approche permet, lorsque la théorie ne le suggère pas a priori, de découvrir la forme de la relation entre la variable dépendante et les variables explicatives.

Halstead *et al* (1997) proposent également d'utiliser la transformation de Box-Cox afin de résoudre le problème du choix de la forme fonctionnelle. Très flexible, l'approche permet au modèle de s'ajuster aux données empiriques. Parmi un ensemble de possibilités, les estimations de la

forme générale suggèrent la forme fonctionnelle la plus appropriée. Bien que les auteurs montrent que, pour le cas de Belcherstown au Massachusetts, la forme log-linéaire soit préférable aux autres et est équivalente à la forme générale, ils mentionnent néanmoins qu'il serait imprudent de généraliser à partir d'un cas particulier parce que chaque application peut nécessiter des transformations particulières. Le choix de la forme fonctionnelle s'avère très spécifique. Ainsi, Fletcher *et al* (2003) suggèrent, pour leur part, la dominance de la forme semi-log sur les alternatives pour leur étude portant sur la région des East Midlands, en Angleterre.

Malgré la puissance de la forme générale de Box-Cox, plusieurs auteurs ont souligné ses lacunes. Maurer *et al* (2004) soulignent la difficulté d'interprétation des coefficients en raison des transformations effectuées. Linneman (1980) mentionne l'impossibilité d'utiliser des variables binaires et des variables prenant des valeurs négatives ou nulles dans une régression de Box-Cox.

Davidson & MacKinnon (1993) montrent que les paramètres de la transformation de Box-Cox affectent les résultats de deux façons différentes : en modifiant la forme fonctionnelle et en changeant le comportement des résidus du modèle. Ils soulignent également que si plusieurs études ont examiné la détermination de la forme fonctionnelle idéale dans les études économétriques, encore peu d'attention a été portée aux propriétés des résidus issus du modèle. Néanmoins, le patron et le comportement des résidus constituent un des concepts clés de la régression par la méthode des moindres carrés ordinaires (MCO). Les transformations sur les variables du modèle ne sont utiles et justifiées que dans la mesure où elles permettent de réduire les résidus et de contrôler leurs propriétés. Fletcher *et al* (2003) mentionnent d'ailleurs que peu de tests sur le comportement des résidus sont rapportés dans les études empiriques.

De plus, dans la spécification de Box-Cox, la transformation des va-

Tableau 1. Statistiques descriptives selon le plan d'échantillonnage : transactions résidentielles unifamiliales - plus de 50 000\$ et moins de 250 000\$

	N	Moyenne	Écart-type	Min	Max
Prix de vente (\$)	19 175	93 748	33 221	50 460	249 000
X ₁ Superficie habitable (m. carrés)	19 175	110,1	34,8	38,3	487,4
X ₂ Taille du terrain (m. carrés)	19 175	659,7	414,3	59,1	9764,7
X ₃ Âge (années)	19 175	18,4	18,0	0,0	311,0
X ₄ Taux de taxe (\$ / 100\$)	19 175	2,41	0,51	1,20	4,13
Vente en 1986	19 175	0,152		0	1
Vente en 1987	19 175	0,029		0	1
Vente en 1988	19 175	0,000		0	1
Vente en 1989	19 175	0,000		0	1
Vente en 1990	19 175	0,164		0	1
Vente en 1991	19 175	0,156		0	1
Vente en 1992	19 175	0,000		0	1
Vente en 1993	19 175	0,155		0	1
Vente en 1994	19 175	0,151		0	1
Vente en 1995	19 175	0,116		0	1
Vente en 1996	19 175	0,078		0	1

riables influence la significativité des paramètres, rendant difficilement justifiable le recours aux statistiques *t*, à l'exception de quelques cas classiques⁴. Ahamada *et al* (2007) suggèrent, dans ce cas, de procéder à l'estimation du modèle de régression avec la transformation de Box-Cox et, à partir des résultats obtenus, de faire le choix de la forme de la relation entre les trois spécifications les plus utilisées, soit la forme log-linéaire (ou log-log), la forme semi-log ou la forme linéaire classique. Ils mentionnent cependant que la présence d'hétéroscédasticité peut aboutir à une estimation du paramètre de transformation inférieure à l'unité, ce qui implique de rejeter l'hypothèse d'une forme fonctionnelle linéaire

Afin de déterminer une forme fonctionnelle dominante sur le territoire d'étude, nous proposons d'estimer les formes fonctionnelles classiques (linéaire, semi-log et log-linéaire), en plus de la forme fonctionnelle générale (Box-Cox), et d'effectuer une panoplie de tests portant sur le comportement prévisionnel des modèles, sur la robustesse des formes fonctionnelles ainsi que sur le comportement des résidus des modèles. La comparaison des résultats des différents tests permet éventuellement d'identifier une forme fonctionnelle qui performe unilatéralement mieux que les autres. Autrement, elle permet de les ordonner en fonction de diverses propriétés plus

ou moins désirables selon l'application visée. Nous cherchons également à identifier les pistes de recherche utiles afin d'améliorer le processus de spécification des équations de prix hédoniques pour la région de Québec ainsi que pour l'ensemble des territoires et domaines d'application.

Méthodologie

La théorie hédonique permet d'exprimer le prix de vente d'une résidence *i* au temps *t*, P_{it} , en fonction des *g* attributs intrinsèques (relatifs à la propriété), X_{git} , et des *h* extrinsèques (relatifs à son environnement), E_{hit} , qui peuvent être résumés dans un vecteur de *k* attributs Z_{kit} . La fonction de prix hédonique est une équation de forme réduite permettant de refléter les influences de l'offre et la demande résidentielle sur la détermination des prix. Dans la plupart des applications empiriques, les formes fonctionnelles utilisées sont de type semi-log ou log-linéaire en raison de leurs nombreux avantages. Ainsi, la première, qui utilise le logarithme du prix comme variable dépendante, arrive à mieux contrôler pour la présence d'hétéroscédasticité alors que la seconde permet en outre d'estimer quasi directement les paramètres d'élasticité.

La forme fonctionnelle paramétrique la plus générale possible est obtenue avec la transformation de Box-Cox. Cette approche est présentée formellement et décortiquée par Hal-

vorsen & Pollakowski (1981). À partir de certaines restrictions sur les paramètres de la forme générale (Équation 1), il est possible de reproduire les formes fonctionnelles les plus utilisées dans la littérature empirique. Formellement, le modèle général est (1), (2), et (3) : voyez Encadré 1. En supposant que le vecteur de paramètres γ_{kl} soit égal à zéro, la forme générale englobe l'ensemble des spécifications linéaires classiques. La première, la forme log-linéaire, est obtenue lorsque les paramètres λ et θ sont égaux à zéro. La forme semi-log est obtenue lorsque le paramètre λ est égal à zéro et θ prend une valeur unitaire. Finalement, la forme linéaire classique est obtenue lorsque les deux paramètres, λ et θ , égalent un.

On obtient des formes fonctionnelles incluant des termes quadratiques et croisés lorsque le vecteur de paramètre γ_k est non nul. Les formes trans-log et semi-log sont obtenues en incluant les termes croisés, tout comme pour la forme linéaire quadratique. Bien que le choix de la forme fonctionnelle puisse s'effectuer à partir d'une spécification emboîtée et d'un test de rapport de vraisemblance, ou avec d'autres statistiques pertinentes (Akaike, 1974), cette étude propose de comparer les résultats d'estimations selon plusieurs critères. Pour ce faire, divers tests statistiques sont effectués à partir des résultats de modèles utilisant des données empiriques.

La première série de tests examine le potentiel prévisionnel des différents modèles. Un des critères repose sur un test de performance/stabilité de la forme fonctionnelle pour chacune des spécifications (test RESET⁵, Ramsey, 1969). Le pouvoir explicatif de chacun des modèles (R^2) est également considéré. La performance prévisionnelle hors échantillon est réalisée avec une partie de l'échantillon (10 %) pour évaluer le modèle estimé sur le complément de l'échantillon. La performance hors échantillon repose sur l'erreur quadratique moyenne (RMSE) ainsi que le calcul d'un pseudo R^2 , défini comme le coefficient de détermination (carré de la corrélation) entre les

Tableau 2. Tests d'égalité des moyennes en fonction de la spécification de l'échantillon

	Transactions valides		Transactions sélectionnées		Test t statistique
	N	Moyenne	N	Moyenne	
Prix de vente (\$)	20 807	92 714	19 175	93 748	-2,70
X ₁ (m. carrés)	20 807	110,6	19 175	110,1	1,57
X ₂ (m. carrés)	20 807	666,9	19 175	659,7	1,67
X ₃ (années)	20 807	19,8	19 175	18,4	7,59
X ₄ (\$ / 100\$)	20 807	2,43	19 175	2,41	3,29
Desservi par l'aqueduc	20 807	0,98	19 175	0,99	-4,42
Vente en 1986	20 807	0,17	19 175	0,15	5,18
Vente en 1987	20 807	0,03	19 175	0,03	2,03
Vente en 1988	20 807	0,00	19 175	0,00	0,00
Vente en 1989	20 807	0,00	19 175	0,00	0,00
Vente en 1990	20 807	0,16	19 175	0,16	-0,90
Vente en 1991	20 807	0,15	19 175	0,16	-1,07
Vente en 1992	20 807	0,00	19 175	0,00	0,00
Vente en 1993	20 807	0,15	19 175	0,15	-1,53
Vente en 1994	20 807	0,15	19 175	0,15	-1,24
Vente en 1995	20 807	0,11	19 175	0,12	-1,06
Vente en 1996	20 807	0,08	19 175	0,08	-0,79

valeurs prédites par le modèle (hors échantillon) et les valeurs réelles.

La seconde série de tests porte sur le comportement des résidus. La présence d'hétéroscédasticité de la variance estimée (White, 1980) et d'autocorrélation spatiale des résidus (Moran, 1950) sont formellement testées. L'hypothèse de normalité des résidus, nécessaire pour assurer la validité des tests statistiques sur les paramètres de modèle, est vérifiée en comparant les valeurs empiriques et théoriques des moments d'ordre 3 et 4 de la distribution des résidus (Jarque & Bera, 1987 ; Bera & Jarque, 1981). Finalement, on applique un test fondé sur une régression auxiliaire permettant d'exprimer les résidus du modèle en fonction des valeurs de la variable dépendante (Brown & Ethridge, 1995). La régression auxiliaire (équation 4), et plus particulièrement le coefficient lié à la variable de prix (transformée), π_1 , apporte un éclairage spécifique sur une possible non-linéarité ou non-stationnarité des paramètres qui peut créer des patrons de sous-estimation ou de surestimation systématique pour certaines tranches de prix. Un coefficient positif et significatif lié à la variable de prix indique que le modèle a tendance à surestimer les propriétés de bas de gamme et à sous-estimer les propriétés de haut de gamme. À l'inverse, un coefficient négatif et significatif indique que le modèle sous-

évalue systématiquement les propriétés les moins chères et surévalue les propriétés les plus coûteuses.

Données

L'approche des MCO est extrêmement sensible aux valeurs extrêmes ou aberrantes. Pour cette raison, la pratique suggère d'utiliser un certain domaine de valeurs pré-identifié afin d'éviter de « contaminer » les résultats avec les valeurs extrêmes. Les valeurs aberrantes peuvent provenir de deux sources : les erreurs de retranscription et de codage et les événements non-observables qui ont une influence réelle sur le processus de négociation entre les vendeurs et les acheteurs. Si la première source peut être contrôlée de façon plutôt simple⁶, il est nettement plus compliqué de contrôler pour la seconde source.

La base de données utilisée décrit les transactions de résidences unifamiliales effectuées sur le territoire de la Ville de Québec, entre 1986 et 1996⁷. Une fois épuré des données manquantes ou erronées et pour un domaine de prix de vente compris entre 50 000\$ et 250 000\$⁸, l'échantillon final s'établit à 19 175 observations. Puisque le but de l'exercice n'est pas d'identifier l'ensemble des variables qui influencent le processus de formation des prix, mais plutôt d'identifier la meilleure forme fonctionnelle, seules

quelques variables sont retenues dans les processus d'estimation (Tableau 1) :

- la superficie habitable (en mètres carrés) ;
- la taille du terrain (en mètres carrés) ;
- l'âge de la résidence ;
- le taux de taxe municipale en vigueur au moment de la vente ;
- le fait d'être desservi par le système d'aqueduc ;
- des variables binaires identifiant l'année de la vente.

La résidence moyenne est âgée de 19 ans, offre une superficie habitable de 110 mètres carrés, possède un terrain de 660 mètres carrés et s'est vendue 93 000\$.

Les écarts entre les statistiques descriptives de cet échantillon et celles de l'ensemble des transactions de la base de données sont relativement faibles. Toutefois, certaines différences sont significatives (Tableau 2). Bien que les principales variables de l'équation de prix (superficie habitable et superficie du terrain) affichent une distribution homogène, on observe des différences significatives dans la distribution du prix de vente⁹, de l'âge moyen des résidences, de leur fardeau fiscal¹⁰ et du nombre de ventes effectuées durant les premières années de la période d'observation. L'écart dans la répartition temporelle peut s'expliquer, en partie, par le processus de sélection, fondé sur la variable dépendante. Puisque les prix de vente sont en valeurs nominales, les chances d'éliminer de l'échantillon des transactions conclues à moins de 50 000\$ sont plus fortes au début de la période. Le phénomène inverse est aussi vrai : les chances d'éliminer de l'échantillon les transactions à plus de 250 000\$ sont plus fortes en fin de période. Une manière alternative de procéder consisterait à établir le critère de sélection sur la valeur réelle des transactions, ajustée pour le coût de la vie (inflation). Cependant, il serait alors nécessaire d'utiliser une mesure adéquate afin de contrôler pour l'évolution des prix dans le secteur ré-

Tableau 3. Performances explicative et prévisionnelle des modèles selon la spécification d'estimation

	Log-linéaire		Semi-log		Linéaire		Box-Cox	
	Statis-tique	Sig	Statis-tique	Sig	Statis-tique	Sig	Statis-tique	Sig
N	19 175		19 175		19 175		19 175	
Forme fonctionnelle (RESET)	152,32 ***		197,09 ***		365,18 ***		4,98 **	
R ²	0,631		0,618		0,627		0,620	
R ² (hors échantillon)	0,615		0,584		0,588		0,603	
écart (valeur absolue)	2,44%		5,40%		6,13%		2,75%	
RMSE ^a	0,189		0,192		20 299		0,0011	
RMSE ^a (hors échantillon)	0,187		0,195		20 883		0,0011	
écart (valeur absolue)	0,73%		1,67%		2,88%		0,66%	
RMSE ^a standardisée	0,608		0,619		0,611		0,616	

Signification : *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$. ^a RMSE divisé par l'écart-type de la variable dépendante (unité de mesure comparable). ^b Erreur quadratique moyenne.

Tableau 4. Tests usuels sur les résidus selon la spécification et la méthode d'estimation : moindres carrés ordinaires.

	Log-linéaire		Semi-log		Linéaire		Box-Cox	
	Statis-tique	Sig	Statis-tique	Sig	Statis-tique	Sig	Statis-tique	Sig
N	19 175		19 175		19 175		19 175	
Hétéroscédasticité ^a	196,01 ***		318,49 ***		1 318,17 ***		37,03 **	
Normalité des erreurs ^b	516,47 ***		774,41 ***		2 920,72 ***		619,35 ***	
Normalité hors échantillon ^b	52,34 ***		126,35 ***		309,73 ***		42,31 ***	
Autocorrélation spatiale ^c	0,1799 ***		0,1799 ***		0,1656 ***		0,1874 ***	

Signification : *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$. ^aWhite. ^bJarque-Berra.

sidentiel et cette tâche déborde les objectifs de cet article.

Résultats

Les modèles sont estimés en appliquant les transformations appropriées sur la variable dépendante et les variables explicatives. Les variables binaires ne subissent aucune transformation¹¹. Le paysage urbain de Québec associe des quartiers densément peuplés à des secteurs de périurbanisation, en passant par des banlieues anciennes et récentes. Pour cette raison, une variable binaire identifiant les résidences desservies par le réseau d'aqueduc est introduite afin de distinguer les secteurs qui ne bénéficient pas de certains services municipaux. Cette variable permet d'isoler, dans une certaine mesure, les effets de localisation dans les zones campagnardes offrant de grands terrains, mais dépourvues de service d'aqueduc et d'égout sanitaire. Ceci permet ainsi

d'obtenir l'effet net de la taille du terrain (desservi versus non desservi) sur la détermination du prix de vente moyen.

Les coefficients estimés selon chaque spécification ne sont pas directement comparables, et ce pour plusieurs raisons. Premièrement, les variables subissent des transformations différentes, mais propres à la forme fonctionnelle sélectionnée. Deuxièmement, étant donné les effets quadratiques proposés dans les formes fonctionnelles plus complexes, les coefficients peuvent changer de signe et d'amplitude. Les seuls modèles comparables, dans une certaine mesure, sont les formes linéaires avec leurs contreparties quadratiques ou généralisées respectives : les formes trans-log versus log-linéaire, les formes semi-log versus semi-log généralisé, les formes linéaire versus linéaire quadratique et les formes Box-Cox versus Box-Cox généralisée.

La comparaison des différents critères de performance des formes fonctionnelles demeure un outil de comparaison utile. En fonction du pouvoir prévisionnel des modèles, les résultats suggèrent une dominance de la forme log-linéaire sur les alternatives classiques lorsque les formes fonctionnelles sont spécifiées selon les spécifications les plus simples, c'est-à-dire sans interaction entre les variables et en excluant les termes quadratiques (Tableau 3). En performance hors-échantillon, le modèle log-linéaire affiche un pouvoir explicatif plus élevé que les autres spécifications, en plus de montrer une meilleure stabilité. Seule ombre au tableau, la statistique de test RESET montre que l'ensemble des modèles ne réussit pas à capter l'essentiel des effets de non-linéarité dans la détermination des prix, suggérant une amélioration de la spécification¹². À une exception près (RESET), la spécification log-linéaire est comparable, et même supérieure en terme de pouvoir prédictif, à l'approche plus générale qu'offre la transformation de Box-Cox.

Les statistiques reliées aux tests effectués sur les résidus des modèles suggèrent également une dominance de la forme log-linéaire (Tableau 4). Les statistiques montrent un comportement moins prononcé sur plusieurs plans, bien que des problèmes soient présents : hétéroscédasticité de la variance, autocorrélation spatiale et non-normalité des termes d'erreurs. Les conclusions sont moins claires lorsqu'on compare les résultats des modèles log-linéaire et de Box-Cox. Si l'approche de Box et Cox montre une variance estimée plus homogène, elle fait cependant état d'une autocorrélation spatiale des résidus légèrement plus importante et possède une statistique de test supérieure quant à la normalité de la distribution des résidus.

En optant pour une spécification plus complexe, les résultats sont un peu moins suggestifs. Si la forme trans-log¹³ possède un pouvoir explicatif plus élevé que les formes semi-log généralisée et que l'approche de Box-Cox, il est légèrement inférieur à celui de la forme linéaire quadratique (Ta-

Tableau 5. Performance prévisionnelle des modèles selon la spécification d'estimation : moindres carrés ordinaires.

	Trans-log		Semi-log généralisé		Linéaire quadratique		Box-Cox généralisé	
	Statis-tique	Sig	Statis-tique	Sig	Statis-tique	Sig	Statis-tique	Sig
N	19 175		19 175		19 175		19 175	
Forme fonctionnelle (RESET)	108,37	***	60,93	***	156,52	***	3,43	
R ²	0,665		0,650		0,669		0,652	
Pseudo-R ² (hors échantillon)	0,643		0,632		0,639		0,636	
écart (valeur absolue)	3,35%		2,76%		4,46%		2,56%	
RMSE ^a	0,180		0,184		19 116		0,0025	
RMSE ^a (hors échantillon)	0,181		0,183		19 544		0,0025	
écart (valeur absolue)	0,57%		0,21%		2,24%		0,46%	
RMSE standardisée	0,579		0,592		0,575		1,363	

Signification : *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$. ^a RMSE divisé par l'écart-type de la variable dépendante (unité de mesure comparable). ^b Erreur quadratique moyenne

Tableau 6. Tests usuels sur les résidus selon la spécification et la méthode d'estimation : moindres carrés ordinaires.

	Trans-log		Semi-log généralisé		Linéaire quadratique		Box-Cox généralisé	
	Statis-tique	Sig	Statis-tique	Sig	Statis-tique	Sig	Statis-tique	Sig
N	19 175		19 175		19 175		19 175	
Hétéroscédasticité ^a	78,82	***	112,40	***	992,68	***	4,87	**
Normalité des erreurs ^b	970,75	***	748,72	***	2923,24	***	715,85	***
Normalité hors échantillon ^b	125,04	***	81,81	***	301,62	***	67,69	**
Autocorrélation spatiale ^c	0,1637	***	0,1644	***	0,1490	***	0,1654	**

Signification : *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$. ^aWhite. ^bJarque-Berra. ^cMoran.

bleau 5). Si on compare la performance hors-échantillon à la performance globale, les approches semi-log et de Box-Cox ont tendance à produire des résultats légèrement plus stables, en termes d'écart, que l'approche log-linéaire. Finalement, le test de RESET montre que, à une exception près, les formes fonctionnelles plus complexes ne parviennent pas à capter l'essentiel de la non-linéarité de la relation, bien qu'elles en diminuent l'effet¹⁴.

En s'attardant aux résultats issus des résidus des différents modèles plus généraux, les formes fonctionnelles utilisant le logarithme du prix comme variable dépendante sont préférables à l'alternative linéaire puisqu'elles permettent de contrôler, en partie, pour les problèmes d'hétéroscédasticité de la variance et surtout pour la non-normalité de la distribution des résidus (Tableau 6). L'approche de Box-Cox généralisée montre des résultats légèrement su-

périeurs à ceux des alternatives utilisant les transformations logarithmiques : le test d'hétéroscédasticité tout comme les tests de normalité des résidus s'avèrent plutôt faibles, autant avec l'échantillon global que hors-échantillon. Seules ombres au tableau, le problème d'autocorrélation spatiale des résidus demeure important et les variables explicatives deviennent plus fortement corrélées (hausse de la multicollinéarité).

La légitimité des variables croisées est justifiée, si on en juge par la significativité des coefficients liés aux variables (Tableau 7 versus Tableau 8). Les variables croisées et quadratiques montrent que la détermination des prix dépend à la fois de leur niveau, mais également de façon simultanée du niveau des autres caractéristiques. Une analyse *ceteris paribus*, sans tenir compte des autres caractéristiques, peut donc s'avérer risquée si ces effets ne sont pas identifiés et contrôlés. Les variables croisées apportent cepen-

dant peu de profondeur supplémentaire au pouvoir explicatif du modèle (un peu plus de trois points de pourcentage) et peu de précision supplémentaire aux modèles (diminution peu importante de l'erreur quadratique moyenne - RMSE). Il est justifié de se demander si une spécification plus complète des différents modèles aurait fait en sorte que les variables croisées et quadratiques auraient influencé leur significativité. Chose certaine, l'omission de variables d'importance dans un modèle a pour effet de biaiser les coefficients obtenus. De plus, si les variables croisées et quadratiques sont fortement significatives, elles entraînent un problème plus marqué de colinéarité entre les variables indépendantes du modèle, ce qui peut fragiliser les résultats. Un autre constat particulièrement intéressant tiré des coefficients estimés repose sur les paramètres liés à la transformation de Box-Cox. Peu importe la spécification retenue, les paramètres sont nettement plus près de zéro que de l'unité, suggérant que le recours à la forme fonctionnelle log-linéaire.

Une autre façon de tester la performance du modèle consiste à effectuer une régression auxiliaire, basée sur les résidus des différents modèles. Cette approche permet de vérifier si le modèle surévalue ou sous-évalue systématiquement certaines valeurs marchandes en fonction de leur gamme de prix. Les résultats montrent que, peu importe la forme fonctionnelle retenue, les modèles font état de patrons d'erreurs systématiques en fonction du prix de vente, bien que le problème soit moins important dans le cas de la forme log-linéaire (Tableau 9). Si le phénomène diminue en importance avec l'ajout de variables croisées et quadratiques, celui-ci demeure néanmoins important. Ces analyses viennent, en quelque sorte, renforcer les résultats du test de RESET. Le problème peut être dû à certains facteurs tels que : l'importance de l'aspect géographique (spatial), l'omission de plusieurs variables significatives, le comportement non-linéaire ou non-stationnaire des paramètres et la nécessité de certaines transformations

Tableau 7. Résultats d'estimation – Formes simples ; moindres carrés ordinaires

	Log-linéaire		Semi-log		Linéaire		Box-Cox	
	Coeff- icient	Sig	Coeff- icient	Sig	Coeff- icient	Sig	Coeff- icient	Sig
X ₁ (m. carrés)	0,6744	***	0,0057	***	644	***	0,0011	***
X ₂ (m. carrés)	0,0664	***	0,0000	***	1,810	***	0,0001	***
X ₃ (années)	-0,0686	***	-0,0040	***	-382	***	-0,0002	***
X ₄ (\$/100\$)	-0,4331	***	-0,1754	***	-18 613	***	-0,0020	***
Desservi par l'aqueduc	0,2221	***	0,1997	***	17 469	***	0,0014	***
1987	0,0554	***	0,0471	***	3 974	***	0,0003	***
1990	0,1331	***	0,1266	***	9 696	***	0,0009	***
1991	0,1652	***	0,1577	***	13 057	***	0,0011	***
1993	0,1947	***	0,1849	***	15 469	***	0,0012	***
1994	0,1790	***	0,1722	***	13 506	***	0,0012	***
1995	0,1654	***	0,1545	***	12 083	***	0,0011	***
1996	0,1739	***	0,1565	***	12 358	***	0,0011	***
Constante	8,0154	***	10,9194	***	45 738	***	2,2095	***
θ							0,2715	***
λ							-0,4477	***
R ²	0,631		0,618		0,627		0,620	
N	19 175		19 175		19 175		19 175	

Signification : *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$.

mathématiques supplémentaires sur les variables explicatives

Discussion

Les tests portant sur la performance prévisionnelle des formes fonctionnelles, sur le comportement des résidus ainsi que sur les coefficients de la forme de Box-Cox nous apprennent certaines leçons quant à la forme fonctionnelle de l'équation de prix hédonique. Premièrement, elles montrent que l'utilisation de la forme log-linéaire est probablement la meilleure alternative linéaire possible selon les paramètres obtenus par la transformation de Box-Cox. Des coefficients près de zéro, mais tout de même significatifs suggèrent la dominance de la forme log-linéaire sur les formes semi-log et linéaire. De plus, règle générale, les statistiques de plusieurs tests avantagent les formes fonctionnelles qui recourent à une transformation logarithmique de la variable dépendante et, le cas échéant, de certaines variables indépendantes.

Deuxièmement, la transformation de Box-Cox ne semble pas constituer le remède à tous les maux, même si les résultats des modèles utilisant les variables croisées et quadratiques suggèrent sa dominance. En plus des lacunes théoriques liées à la transforma-

tion de Box-Cox (Davidson & MacKinnon, 1993), l'interprétation des coefficients devient plus complexe, tout comme le calcul des effets marginaux. Bien que le test RESET demeure intéressant pour vérifier la pertinence de la spécification retenue, son utilisation est fragilisée lorsqu'il existe une forte corrélation entre les variables explicatives. La significativité du I de Moran (autocorrélation spatiale globale) et du coefficient associé à la variable prix dans la régression auxiliaire montre la présence de certains problèmes dans le comportement des résidus, et ce, de façon plus prononcée qu'avec l'approche log-linéaire ou trans-log.

Malgré la spécification plutôt simpliste retenue dans la présente analyse, le test de RESET et le test issu de la régression auxiliaire permettent de mettre en lumière une certaine complexité dans la relation empirique reliant le prix de vente aux attributs d'une résidence. Si les variables croisées et quadratiques incluses dans les formes fonctionnelles généralisées permettent de tenir compte d'une certaine forme de non-linéarité dans le processus de détermination des valeurs marchandes, le problème n'en demeure pas moins présent. La question de détermination des prix hédoniques aurait donc avantage à être analysée en suggérant certaines trans-

formations sur les variables continues du modèle de prix hédonique. L'interaction des variables, les transformations quadratiques, mais aussi certaines transformations mathématiques, telles que la transformation gamma (Des Rosiers et al 1996 ; 2001), pourraient apporter de la profondeur aux équations de prix existantes et ainsi assurer une certaine stabilité des coefficients de la forme fonctionnelle.

Un autre problème souvent rencontré dans la littérature pour l'estimation des modèles de prix hédoniques, mais souvent peu explicité, repose sur la sélection d'un sous-échantillon en fonction du domaine de valeur de la variable dépendante. Limiter l'analyse à un certain domaine de valeurs introduit un biais dans les coefficients estimés puisque la distribution de la variable dépendante est tronquée et, par conséquent, les paramètres de la distribution (moyenne et variance) peuvent changer de façon marquée. Les paramètres obtenus en limitant l'analyse à un certain domaine de valeurs sont localement valables, sans l'être nécessairement de façon globale. Cependant, même en ajustant les modèles pour corriger la troncature de la distribution de la variable dépendante, autant à droite qu'à gauche (Tobin, 1958 ; Amemiya, 1973 ; Goldberger, 1983), les conclusions sont semblables à celles tirées précédemment. Seule différence notable : les coefficients liés à la variable de prix dans les régressions auxiliaires sont plus faibles, supportant ainsi, en partie, la thèse voulant que la méthode d'estimation limitée à un domaine de valeurs introduise certains biais. À ce sujet, un examen plus attentif de la sélection des valeurs de la variable dépendante s'avère important ; on pourrait par exemple suggérer de choisir le domaine de valeurs sur la base d'une analyse fine de la distribution de fréquence des prix de façon à limiter les exclusions aux observations les plus excentriques.

À la lumière de ces constats, une question émerge quant à la méthode d'estimation à privilégier pour spécifier l'équation de prix. Le problème de surestimation et de sous-estimation systématique du modèle de prix hé-

Tableau 8. Résultats d'estimation – Formes généralisées ; moindres carrés ordinaires

	Trans-log		Semi-log		Quadratique		Box-Cox	
	Coeff- icient	Sig	Coeff- icient	Sig	Coeff- icient	Sig	Coeff- icient	Sig
X ₁ (m. carrés)	-1,2291 ***		0,0089 ***		1 019 ***		0,1053 ***	
X ₂ (m. carrés)	0,5546 ***		0,0001 ***		18 ***		-0,0488 ***	
X ₃ (années)	0,3601 ***		0,0001		107 *		-0,0220 ***	
X ₄ (\$/100\$)	0,5680 ***		-0,5219 ***		-47 109 ***		0,0141 ***	
X ₁ ²	0,2136 ***		0,0000 ***		0,101		-0,1134 ***	
X ₁ * X ₂	-0,0004		0,0000 ***		-0,050 ***		-0,0030	
X ₁ * X ₃	-0,0166 ***		0,0000 ***		-2,411 ***		0,0015	
X ₁ * X ₄	-0,0860 ***		-0,0002 *		-137 ***		-0,0019	
X ₂ ²	-0,0235 ***		0,0000		0,00		0,0258 ***	
X ₂ * X ₃	-0,0262 ***		0,0000 *		0,025 *		0,0101 ***	
X ₂ * X ₄	-0,1295 ***		0,0000 *		-3,659 ***		0,0439 ***	
X ₃ ²	-0,0192 ***		0,0000 ***		2,131 ***		0,0042 ***	
X ₃ * X ₄	-0,1195 ***		-0,0021 ***		-150 ***		0,0146 ***	
X ₄ ²	0,3626 ***		0,0868 ***		10 296 ***		-0,0229 ***	
Desservi par l'aqueduc	0,2133 ***		0,2170 ***		19 665 ***		0,0030 ***	
1987	0,0609 ***		0,0508 ***		4 086 ***		0,0009 ***	
1990	0,1679 ***		0,1639 ***		14 045 ***		0,0024 ***	
1991	0,1957 ***		0,1937 ***		17 034 ***		0,0028 ***	
1993	0,2339 ***		0,2273 ***		20 208 ***		0,0033 ***	
1994	0,2240 ***		0,2135 ***		18 336 ***		0,0032 ***	
1995	0,2093 ***		0,1990 ***		17 130 ***		0,0030 ***	
1996	0,2145 ***		0,2009 ***		17 281 ***		0,0030 ***	
Constante	9,9176 ***		10,9902 ***		39 523 ***		2,6548 ***	
θ							-0,2281 **	
λ							-0,3746 ***	
R ²	0,665		0,650		0,669		0,652	
N	19 175		19 175		19 175		19 175	

Signification : *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$

donique implique que les coefficients du modèle deviennent potentiellement non stationnaires, alors que le modèle linéaire usuel suppose que la relation est globalement stable. La non-stationnarité des paramètres incorpore, implicitement, une partie du phénomène d'autocorrélation spatiale des résidus. L'absence de stationnarité des coefficients est modélisable en introduisant des complexités dans la détermination des coefficients en fonction de certains segments de prix. Le modèle de base peut être ajusté en utilisant la méthode d'expansion des coefficients (Casetti, 1972 ; 1997) ou toute autre technique appropriée. L'estimation de tels modèles permet de déterminer si les paramètres associés aux caractéristiques varient en fonction du prix de vente.

La méthode d'estimation du modèle de prix hédoniques, potentiellement paramétrique, devrait permettre aux coefficients de varier selon certains segments de prix tout en élimi-

nant les problèmes reliés aux valeurs extrêmes. À ce sujet, la régression par quantile (Koenker & Bassett, 1978 ; Koenker, 2005) apparaît comme une approche intéressante et naturelle : elle permet de considérer l'ensemble des observations sans être trop sensible aux valeurs extrêmes, tout en permettant une certaine dérive (ou non-stationnarité) des paramètres. Les résultats d'estimations permettent d'obtenir un ensemble des coefficients lié à une variable explicative en fonction de la distribution de la variable dépendante. Cette approche permet d'étudier l'hypothèse de stabilité des prix hédoniques et d'identifier l'alternative, soit des processus non stationnaires.

Conclusion

Cet article a cherché à déterminer la meilleure forme fonctionnelle paramétrique possible dans l'évaluation du processus de formation des prix résidentiels pour le territoire de la ville de

Québec. De façon générale, il ressort que la forme log-linéaire est supérieure aux alternatives qu'offrent les formes linéaires et semi-logarithmiques. Cette conclusion est appuyée par différents tests sur les performances prévisionnelles et les résidus des modèles. La forme log-linéaire traditionnelle peut également être bonifiée en introduisant des effets plus complexes par le recours aux variables croisées ou quadratiques. L'introduction de variables plus complexes permet d'améliorer les performances prédictives des modèles, tout en diminuant légèrement la présence d'autocorrélation spatiale globale dans les résidus.¹⁵

Les résultats vont dans le même sens si on utilise une approche permettant de tenir compte de la sélection de l'échantillon sur un certain domaine de valeur de la variable dépendante. La troncature de l'échantillon a pour effet de modifier volontairement les paramètres de la distribution de la variable dépendante et de rendre les résultats valables localement, sans être généralisable globalement. Néanmoins, les résultats montrent que cette méthode de correction, quoiqu'ayant une influence sur l'estimation des paramètres, n'améliore pas nécessairement le comportement prévisionnel des modèles ainsi que le comportement des résidus des équations.

L'exercice permet également d'exprimer certains bémols sur l'utilisation de la forme fonctionnelle log-linéaire. Le constat le plus intéressant provient d'une analyse de régression auxiliaire sur les résidus des formes fonctionnelles. Cette analyse montre qu'il existe un patron de surestimation systématique pour les prix des résidences en deçà de la moyenne et de sous-estimation systématique des prix de vente pour les résidences dont les prix sont situés au-delà de la moyenne. Ce comportement peut être attribuable à trois effets, soit la localisation et les externalités associées qui ne sont pas prises en compte dans le modèle (reflétée dans le calcul des I de Moran), la non-stationnarité des coefficients (prix hédoniques) en fonction de la segmentation du marché sur la

Tableau 9. Biais détectable dans les termes d'erreurs – Tests sur des régressions auxiliaires

	Log-linéaire		Semi-log		Linéaire		Box-Cox		
	Coeff- icient	Sig ifiant	Coeff- icient	Sig ifiant	Coeff- icient	Sig ifiant	Coeff- icient	Sig ifiant	
Formes simples									
Prix de vente	0,3694 ***		0,3824 ***		0,3731 ***		0,3797 ***		
Constante	-4,2096 ***		-4,3584 ***		-34 981 ***		-0,8429 **		
R ²	0,369		0,382		0,370		0,380		
N	19 175		19 175		19 175		19 175		
Log-vraisemblance	9 207		9 074		-212 908		107 650		
Critère d'Akaike	-18 410		-18 144		425 820		-215 296		
Formes généralisées									
			Trans- -log		Semi-log généralisé		Quad- ratique		Box-Cox généralisé
			Coeff- icient	Sig ifiant	Coeff- icient	Sig ifiant	Coeff- icient	Sig ifiant	Coeff- icient
Prix de vente	0,3347 ***		0,3498 ***		0,3307 ***		0,3478 **		
Constante	-3,8145 ***		-3,9860 ***		-31 005 ***		-0,9155 ***		
R ²	0,335		0,350		0,331		0,348		
N	19 175		19 175		19 175		19 175		
Log-vraisemblance	9 639		9 436		-212 379		91 956		
Critère d'Akaike	-19 273		-18 869		424 762		-183 908		
Signification : *** p<0,001; ** p<0,01; * p<0,05									

base du prix de vente, et la non-stationnarité en fonction de variations spatiales effectives des coefficients non exprimées par le modèle (autocorrélation spatiale locale).

Les deux dernières situations auraient pour effet de rendre la fonction de prix hédonique instable dans les paramètres en fonction du domaine de valeur de la variable dépendante ou de la localisation de la propriété transigée. La régression par quantile s'avère une approche intéressante permettant de vérifier l'hypothèse de stationnarité des prix implicites (ou des contributions marginales) des caractéristiques et de dresser le portrait de l'évolution des paramètres en fonction du prix de vente. À ce sujet, nous proposons d'évaluer cette possibilité dans des recherches futures. Cette approche s'apparente également à une méthode de segmentation du marché immobilier, mais sur la base de la variable dépendante plutôt que sur la localisation géographique. Implicitement, la segmentation par les prix (par exemple, par quartiers) permettrait de contrôler pour une certaine dérive spatiale des coefficients. Après

tout, les quartiers sont relativement homogènes et les processus de détermination des prix sont interdépendants des facteurs de choix de localisation résidentielle. Enfin, les perturbations liées à l'autocorrélation spatiale locale peuvent être pris en charge (contrôlés) par les méthodes autorégressives.

Références

- Ahamada, I, et al. 2007. Prix des logements et autocorrélation spatiale : Une approche semi-paramétrique. *Économie publique* 20(1), 131-145.
- Akaike, H. 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19(6), 716-723.
- Amemiya, T. 1973. Regression Analysis when the Dependent Variable is Truncated Normal. *Econometrica* 41, 997-1016.
- Anselin, L. 1995. Local Indicators of Spatial Association – LISA. *Geographical Analysis* 27(2), 93-115.
- Anselin, L, & Florax, R. 1995. Introduction. Dans L. Anselin and R. Florax (Eds.), *New Directions in Spatial Econometrics*, Berlin : Springer-Verlag : 3-18.
- Bera, AK, & Jarque, CM. 1981. *An efficient Large Sample Test for Normality of Observations and Regression Residuals*, Australian National University Working Papers in Econometrics, no. 40, Australian National University, Canberra. 31 pages.
- Blackley, P, et al. 1984. Box-Cox Estimation of Hedonic Models : How Serious is the Iterative OLS Variance Bias ? *Review of Economics and Statistics* 66(2), 348-353.
- Box, GEP, & Cox, DR. 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B (Methodological)* 26(2), 211-252.
- Brown, JE, & Ethridge, DE. 1995. Functional Form Model Specification : An Application to Hedonic Pricing. *Agricultural and Resource Economics Review*, October : 166-173.
- Can, A. 1992. Specification and Estimation of Hedonic Housing Price Models. *Regional Science and Urban Economics* 22(3), 453-474.
- Casetti, E. 1997. The Expansion Method, Mathematical Modeling and Spatial Econometrics. *International Regional Science Review* 20(1-2), 9-33.
- Casetti, E. 1972. Generating Models by the Expansion Method : Applications to Geographical Research. *Geographical Analysis* 4(1), 81-91.
- Christensen, LR, et al. 1973. Transcendental Logarithmic Production Frontiers. *Review of Economics and Statistics* 55(1), 28-45.
- Christensen, LR, et al. 1971. Conjugate Duality and the Transcendental Logarithmic Production Function. *Econometrica* 39, 255-256.

- Davidson, R, & MacKinnon, JG. 1993. *Estimation and Inference in Econometrics*, Oxford University Press, New York, New-York.
- Des Rosiers, F, et al. 2001. Size and Proximity Effects of Primary Schools on Surrounding House Values. *Journal of Property Research* 18(2), 1-20.
- Des Rosiers, F, et al. 1996. Shopping Centers and House Values : An Empirical Investigation, *Journal of Property Valuation and Investment* 14(4). 41-63.
- Dubé, J, et al. 2011. Impact de la segmentation spatiale sur le de la forme fonctionnelle pour la modélisation hédonique. *Revue d'économie régionale et urbaine* 1, 9-38.
- Fletcher, M, et al. 2003. Comparing Hedonic Models for Estimating and Forecasting House Prices. *Property Management* 22(3/4), 189-200.
- Goldberger, AS. 1983. Abnormal Selection Bias, dans Karlin, S, Amemiya, T. & Goodman, L.A. (Eds.) *Studies in Econometrics, Time Series, and Multivariate Statistics*, Academic Press, New York : 67-84.
- Goodman, AC. 1978. Hedonic Prices, Price Indices and Housing Markets. *Journal of Urban Economics* 5(4), 471-484.
- Goodman, AC, & Thibodeau, TG. 1995. Age-Related Heteroskedasticity in Hedonic House Price Equations. *Journal of Housing Research* 6(1), 25-42.
- Halstead, JM, et al. 1997. On the Issue of Functional Form Choice in Hedonic Price Functions : Further Evidence. *Environmental Management* 21(5), 759-765.
- Halvorsen, R, & Pollakowski, HO. 1981. Choice of Functional Form for Hedonic Price Equations, *Journal of Urban Economics* 10(1), 37-49.
- Jarque, CM, & Bera, AK. 1987. A Test for Normality of Observations and Regression Residuals. *International Statistics Review* 55, 163-177.
- Koenker, R. 2005. *Quantile Regression*, Econometric Society Monographs vol. 38, Cambridge University Press, New York.
- Koenker, R. & Bassett, G. 1978. Regression Quantiles. *Econometrica* 46, 33-50.
- Lau, J. 1974. Applications of Duality Theory : A Comment, dans Intriligator, M.D. (Eds.) *Frontiers of Quantitative Economics*, vol. 2, North-Holland, Amsterdam. 176-199.
- Linneman, P. 1980. Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing Market. *Journal of Urban Economics* 8, 47-68.
- Maurer, R, et al. 2004. Hedonic Price Indices for the Paris Housing Market. *Allgemeines Statistisches Archiv* 88, 303-326.
- Moran, PAP. 1950. Notes on Continuous Stochastic Phenomena. *Biometrika* 37(1), 17-23.
- Ramsey, JB. 1969. Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the American Statistical Association* 69, 935-939.
- Rosen, S. 1974. Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition. *Journal of Political Economy* 82(1), 34-55.
- Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica* 26, 24-36.
- White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48, 817-838.
- Wooldridge, J.M. 2000. *Introductory Econometrics : A Modern Approach*, USA : South-Western College Publishing.
- ¹ $\ln Y = f[\ln X_i]$.
- ² $\ln Y = f[X_i]$.
- ³ Christensen et al (1971, 1973) et Lau (1974).
- ⁴ Essentiellement, la statistique t s'applique à la forme linéaire – soit en l'absence de toute transformation - ou à la fonction exponentielle qui requiert une transformation logarithmique sur la variable dépendante (modèle semi-log).
- ⁵ RESET : Regression specification error test (Wooldridge, 2000 : p. 281).
- ⁶ En éliminant les observations aberrantes.
- ⁷ Dubé et al (2011) montrent que pour la décennie, les contributions marginales des attributs sont constantes dans le temps. Par contre, la base de données ne contient aucune information sur les ventes effectuées dans les deux derniers trimestres de 1986 ainsi que pour les années 1987, 1988 et 1992.
- ⁸ L'élimination de certaines observations du processus d'estimation est susceptible d'engendrer des biais importants sur les coefficients, en tronquant la moyenne et en diminuant la variance de la distribution de la variable dépendante. Ce problème sera abordé dans la section discussion des résultats.
- ⁹ Puisque le processus de sélection s'effectue sur cette base.
- ¹⁰ Cette variable a une dimension spatiale et temporelle puisque le taux de taxe varie d'une année à l'autre en fonction, notamment, des besoins de financement de chaque municipalité.
- ¹¹ Les coefficients associés aux années de vente permettent de reconstruire un indice de l'évolution des prix sur la décennie observée et tiennent compte de la valeur nominale, plutôt que réelle, des transactions.
- ¹² Cet effet peut également être attribuable à la spécification plutôt simpliste de l'équation de prix.
- ¹³ Une généralisation de l'approche log-linéaire faisant intervenir une série d'interactions (termes croisés) entre les variables explicatives.
- ¹⁴ Évidemment, la statistique s'interprète en fonction de la spécification plutôt simple de l'équation de prix. Il est aussi important de mentionner que cette conclusion doit être modérée puisque les modèles plus complexes introduisent une certaine colinéarité entre les variables explicatives, ce qui peut fragiliser la statistique de test et ainsi biaiser les résultats.
- ¹⁵ Pour ce travail, nous n'avons pas étudié les phénomènes d'autocorrélation spatiale locale (Anselin & Florax, 1995 ; Anselin, 1995) avec une méthode de type LISA (*local indicator of spatial autocorrelation*), ce qui constitue une étape ultérieure de cette recherche.