



Quelque part dans le temps :
contribution de la dimension temporelle dans l'analyse de microdonnées en coupes
transversales empilées

Thèse présentée
dans le cadre du programme de doctorat en Développement régional
en vue de l'obtention du grade de Philosophiae doctor (Ph.D.)

PAR
© NICOLAS DEVAUX

Avril 2017

Composition du jury :

Yann Fournis, président du jury, Université du Québec à Rimouski

François Des Rosiers, examinateur externe, Université Laval

Thierno Diallo, examinateur externe, Université du Québec à Chicoutimi

Mario Handfield, directeur de recherche, Université du Québec à Rimouski

Jean Dubé, codirecteur de recherche, Université Laval

Dépôt initial le 29 août 2016

Dépôt final le [11 avril 2017]

UNIVERSITÉ DU QUÉBEC À RIMOUSKI
Service de la bibliothèque

Avertissement

La diffusion de ce mémoire ou de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire « *Autorisation de reproduire et de diffuser un rapport, un mémoire ou une thèse* ». En signant ce formulaire, l'auteur concède à l'Université du Québec à Rimouski une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de son travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, l'auteur autorise l'Université du Québec à Rimouski à reproduire, diffuser, prêter, distribuer ou vendre des copies de son travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de la part de l'auteur à ses droits moraux ni à ses droits de propriété intellectuelle. Sauf entente contraire, l'auteur conserve la liberté de diffuser et de commercialiser ou non ce travail dont il possède un exemplaire.

À ma mère Sophie.

AVANT-PROPOS

Désireux d'explorer de nouveaux horizons, l'entreprise de cette aventure doctorale est comparable à l'exploration d'un nouveau monde (pour moi du moins). Tout débuta en 2012-2013 avec la rencontre de Jean Dubé dont les travaux et les méthodes m'étaient jusqu'alors inconnus. Ce dernier m'a pourtant introduit au monde de l'analyse spatiale en m'invitant à contribuer à un projet de recherche en cours : « étudier l'impact de l'omission de la dimension temporelle dans l'analyse de données en coupes transversales empilées ». La tâche n'avait rien de commune mais le défi était trop stimulant pour laisser passer cette chance.

L'analyse régionale ou territoriale représente un vaste monde. Tandis que l'espace peut être identifié comme un élément catalyseur de la naissance de la science régionale, la dimension temporelle n'en est pas moins centrale. La mise à disposition croissante de microdonnées (données désagrégées à l'échelle individuelle) ouvre de nouvelles avenues mais pose certains défis notamment liés à l'inclusion de la dimension temporelle dans les méthodes économétriques spatiales. En effet, la formalisation des processus de dépendance en contexte spatiotemporel est primordiale pour l'estimation des effets marginaux totaux et, par conséquent, pour la validité des résultats de recherche. L'adaptation des méthodes économétriques spatiales à ces données constitue un champ de recherche important dans la compréhension des dynamiques régionales ou encore l'analyse de politiques publiques.

La thèse concrétise plusieurs années de travail durant lesquelles j'aurai découvert un champ de recherche stimulant, développé des compétences pertinentes et vécu des expériences enrichissantes tant sur le plan professionnel qu'humain. Tout cela n'aura pas pu être possible sans la contribution de plusieurs personnes.

Le temps... Il est tout ce que nous avons. Je ne saurais trop remercier Jean Dubé de m'en avoir consacré autant, d'avoir cru en moi et pour son soutien indéfectible. Son apport a certes été financier par l'intermédiaire de fonds de recherche mais il a aussi été bien au-delà de cela. On s'est « enfargé les pieds dans les fleurs du tapis » ici et là, mais le « Coach Dubé » a transformé ce parcours doctoral en une aventure qui valait le coup d'être vécue. Grâce à son enthousiasme et la liberté qu'il m'a donnée, j'ai eu l'opportunité de contribuer activement à des projets de recherche en lien direct et en parallèle de cette thèse, de rédiger des articles scientifiques et de participer à plusieurs colloques d'envergure internationale.

Premier auteur de tous les articles présentés dans cette thèse, ma démarche de formation a été marquée par des travaux d'équipe avec des chercheurs d'horizons variés. À ce titre, je tiens à remercier chacun des co-auteurs impliqués dans ce processus : Diego Legros, Philippe Apparicio et Etienne Berthold. Je tiens d'ailleurs à souligner particulièrement la contribution de Diego Legros à mon cheminement. Ses explications, son support et son accueil à Dijon sont inestimables. Un grand merci également à « Mamie Dijon » qui a grandement facilité mon séjour. Je désire ensuite mentionner et remercier Mario Handfield dont le soutien a commencé il y a déjà longtemps, avant même le début de cette thèse.

Sur un plan plus personnel, je me dois de remercier mes proches et particulièrement ma mère Sophie, elle qui m'a accompagné et encouragé depuis toujours et sans jamais fixer de limite à mes ambitions. Sans elle, tant d'expériences n'auraient pas pu être vécues. Je lui en suis à jamais reconnaissant.

Une partie du travail et la diffusion des résultats ont pu être réalisées grâce à des subventions provenant des Fonds de recherche du Québec – Société et culture (FRQSC). Sans cette contribution, la recherche n'aurait pas pu se concrétiser. L'obtention de bourses doctorales d'excellence en recherche ou de mobilité provenant de la Fondation de l'UQAR

et du Centre de recherche en développement territorial (CRDT) ont aussi été des contributions significatives.

*« Through fiction we saw the birth of futures yet to come
Yet in fiction lay the bones ugly in their nakedness
Yet under this mortal sun, we cannot hide ourselves. »*

(In Fiction, *Panopticon*, Isis, 2004)

RÉSUMÉ

Dans le contexte d'une disponibilité croissante des microdonnées (données désagrégées à l'échelle individuelle) non répétées, il est commun de constater l'application de méthodes économétriques spatiales en ignorant la dimension temporelle. Les relations spatiales se caractérisent par leur nature multidirectionnelle, tandis que les relations dans le temps respectent une logique unidirectionnelle. L'hypothèse principale de la recherche suggère qu'en omettant la dimension temporelle le chercheur postule la simultanéité des observations, surestimant ainsi les relations spatiales. En conséquence, les statistiques spatiales estimées sont faussées nuisant au contrôle de la dépendance temporelle et aux résultats de recherche. La thèse vise à i) évaluer les conséquences d'une telle omission sur les statistiques de détection et de correction de la dépendance spatiale et ii) à présenter des approches de modélisation basées sur un estimateur de différence-en-différences (DID) permettant un contrôle des effets spatiaux et temporels.

La thèse peut être décomposée en six sections. La première section propose une introduction générale présentant le contexte ainsi que la problématique de la recherche. La seconde et la troisième section discutent respectivement du cadre théorique et méthodologique de la recherche. La notion d'espace dans la science régionale ainsi que certains fondements de l'économétrie spatiale et de la méta-analyse y sont notamment présentés. À partir de données de transactions immobilières, la quatrième section regroupe les deux premiers articles de la thèse misant sur la méta-analyse afin d'évaluer l'effet de l'omission de la dépendance temporelle sur l'analyse de données en coupes transversales empilées. La cinquième section regroupe les deux articles suivants proposant des applications concrètes permettant de contrôler en large partie l'effet de variation spatiale pour se concentrer sur la dimension temporelle, et l'estimation d'un modèle DID. Finalement, une conclusion générale présente une synthèse des résultats.

Les conclusions de la recherche soulignent la contribution de la dimension temporelle aux analyses. En effet, si les observations sont situées dans l'espace, elles le sont également dans le temps. L'omission de la dépendance temporelle affecte la mesure des effets marginaux et de débordement. Les méthodes économétriques spatiotemporelles devraient ainsi être priorisées au risque d'invalider les résultats d'études. La portée de ces résultats s'étend à tous les champs de recherche régionale, urbaine ou territoriale impliquant l'analyse quantitative de microdonnées non répétées.

Mots clés : Économétrie spatiale; Autocorrélation spatiale; Modélisation spatiotemporelle; Modèle de prix hédonique; Science régionale; Microdonnées

ABSTRACT

Unrepeatedly collected and geolocated micro data (disaggregated data at the individual level) are increasingly available. Despite the presence of a temporal dimension, it is not uncommon for these data to be analyzed using a strictly spatial approach. The challenge lies in the different nature of the spatial (multidirectional) and temporal (unidirectional) relations. The main assumption of the research suggests that by omitting the temporal dimension the researcher postulates the simultaneous observation of the data, thus overestimating the spatial relationships. Accordingly, the estimated spatial statistics are affected invalidating study results. The thesis aims to i) evaluate the consequences of such an omission on spatial dependency detection and correction statistics and ii) to present a modeling approach based on a difference-in-differences (DID) estimator controlling for spatial and temporal effects.

The thesis can be divided into six sections. The first section provides a general introduction presenting the general context and motivation of the research. The second and third sections respectively discuss the theoretical and methodological framework of the research. The notion of space in regional science and some foundations of spatial econometrics and the meta-analysis methods are presented. Using real estate data, the fourth section contains the first two papers of the thesis focusing on the meta-analysis to assess the effect of the omission of the temporal dimension in cross-sectional micro-data pooled over time. The fifth section provides the following two articles offering practical applications based on a DID estimator controlling for spatial variation effects to focus on the temporal dimension. Finally, a general conclusion summarizes the results.

Overall, the research findings highlight the contribution of the temporal dimension to the analysis. Indeed, while the observations are located in space, they also are situated in time. The omission of the temporal dimension will adversely affect marginal and spillover effect measures. Spatiotemporal econometric methods should be prioritized for valid study results. These findings can be extended to all regional, urban or territorial quantitative analysis involving spatial micro data pooled over time.

Keywords: Spatial econometrics; Spatial autocorrelation; Spatiotemporal modelling; Hedonic pricing model; Regional science; Microdata

TABLE DES MATIÈRES

AVANT-PROPOS.....	ix
RÉSUMÉ.....	xiii
ABSTRACT.....	xv
TABLE DES MATIÈRES.....	xvii
LISTE DES TABLEAUX.....	xxiii
LISTE DES FIGURES.....	xxv
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES.....	xxvii
LISTE DES SYMBOLES.....	xxix
INTRODUCTION.....	1
CHAPITRE 1 SCIENCE REGIONALE ET ECONOMETRIE SPATIALE.....	7
1.1 ÉMERGENCE DE LA SCIENCE REGIONALE.....	7
1.1.1 Un vaste champ de recherches.....	7
1.2 ESPACE ET LOCALISATION.....	11
1.2.1 La notion d'espace dans l'analyse régionale.....	11
1.2.2 Définir l'espace géographique.....	14
1.3 CONTRIBUTION DE L'ECONOMETRIE SPATIALE.....	24
1.3.1 Développement de l'économétrie spatiale.....	24
1.3.2 Phénomène de dépendance spatiale.....	29
1.4 PROBLEMATIQUE.....	32
1.4.1 Échelle d'analyse et structure des données.....	32
1.4.2 Question et objectifs de recherche.....	37
CHAPITRE 2 CADRE THÉORIQUE ET EMPIRIQUE.....	39

2.1	CADRE THEORIQUE	40
2.1.1	Théorie hédonique	40
2.2	CADRE EMPIRIQUE	44
2.2.1	Le modèle de régression linéaire multiple (RLM) standard, MCO.....	44
2.2.2	Les défis du processus générateur de données (PGD).....	47
2.2.3	Intégration de l'espace et du temps au PGD.....	50
2.3	DE LA THEORIE A LA PRATIQUE : SCHEMA CONCEPTUEL	53
CHAPITRE 3 MÉTHODOLOGIE.....		87
3.1	MODELISATION DES PROCESSUS SPATIAUX.....	88
3.1.1	La matrice de pondérations spatiales (S).....	89
3.1.2	Démarches de modélisation et tests de détection des processus spatiaux	92
3.1.3	Correction de la dépendance spatiale : les effets fixes et leurs limites.....	98
3.1.4	Correction de la dépendance spatiale : les modèles autorégressifs spatiaux	99
3.2	MODELISATION DES PROCESSUS SPATIOTEMPORELS	102
3.2.1	Conjuguer temps et espace	102
3.2.2	La matrice de pondérations spatiotemporelles	104
3.3	META-ANALYSE ET META-REGRESSION	108
3.3.1	Définition.....	108
3.3.2	Méta-régression et effet de taille (Effect Size).....	110
3.3.3	Avantages et limites de la méta-analyse.....	112
CHAPITRE 4 REVUE DE LA LITTÉRATURE.....		117
4.1	ÉTUDES UTILISANT UNE APPROCHE EN COUPES TRANSVERSALES.....	118
4.2	ÉTUDES UTILISANT UNE APPROCHE SPATIOTEMPORELLE	123
CHAPITRE 5 L'INFLUENCE DU TEMPS SUR LA DÉPENDANCE SPATIALE : UNE MÉTA-ANALYSE BASÉE SUR L'UTILISATION DE LA MODÉLISATION HÉDONIQUE EN IMMOBILIER.....		133

5.1	PRESENTATION DE L'ARTICLE	133
5.2	ABOUT THE INFLUENCE OF TIME ON SPATIAL DEPENDENCE: A META-ANALYSIS USING REAL ESTATE HEDONIC PRICING MODELS.....	135
5.2.1	Abstract.....	135
5.2.2	Introduction.....	135
5.2.3	Spatial and temporal dimensions in real estate data	139
5.2.4	Dealing with spatial autocorrelation: detection and correction	144
5.2.5	Methodology: The meta-analysis and methodological framework	147
5.2.6	Data collection.....	150
5.2.7	Descriptive statistics	156
5.2.8	Estimation results: Spatial dependence detection, Moran's <i>I</i> index.....	161
5.2.9	Estimation results: Spatial dependence correction	165
5.2.10	Discussion.....	169
5.2.11	Conclusion.....	171
 CHAPITRE 6 QUAND TOBLER RENCONTRE HAGERSTRAAND : L'IMPORTANCE DES DIMENSIONS SPATIALES ET TEMPORELLE DANS L'ANALYSE STATISTIQUE		
6.1	PRESENTATION DE L'ARTICLE	175
6.2	WHEN TOBLER MEETS HÄGERSTRAAND: ABOUT THE IMPORTANCE OF SPATIAL AND TEMPORAL DIMENSIONS IN STATISTICAL ANALYSIS	177
6.2.1	Abstract.....	177
6.2.2	Introduction.....	177
6.2.3	Spatiotemporal data and hedonic price modeling: a brief literature review	179
6.2.4	Methodology.....	182
6.2.5	Data and Sampling Procedure	185
6.2.6	Analysis results.....	191

6.2.7	Discussion and Conclusion.....	197
CHAPITRE 7 IMPACT DES POLITIQUES DE PATRIOMONIALISATION SUR LES VALEURS RÉSIDENTIELLES : LE CAS DU VIEUX-QUÉBEC.....		
199		
7.1	PRÉSENTATION DE L'ARTICLE.....	199
7.2	THE IMPACT OF HERITAGE POLICIES ON RESIDENTIAL PROPERTY VALUES: THE CASE OF THE OLD CITY OF QUÉBEC.....	201
7.2.1	Abstract.....	201
7.2.2	Introduction.....	202
7.2.3	The old city of Québec historic district and the reorganization of D'Auteuil Street as heritage policies.....	204
7.2.4	Heritage Policies and Hedonic Modeling of Residential Property Values in Historic Districts.....	212
7.2.5	Methodological Framework.....	214
7.2.6	Empirical application and transaction data.....	220
7.2.7	Main findings and estimation results.....	222
7.2.8	Conclusion.....	229
CHAPITRE 8 EFFET D'ANTICIPATION ET POST-CONSTRUCTION DE L'EXTENSION D'UNE LIGNE DE MÉTRO SUR LES VALEURS RÉSIDENTIELLES : LE CAS DE LAVAL (CANADA), 1995-2013.....		
233		
8.1	PRÉSENTATION DE L'ARTICLE.....	233
8.2	ANTICIPATION AND POST-CONSTRUCTION IMPACT OF THE METRO EXTENSION ON RESIDENTIAL VALUES: THE CASE OF LAVAL (CANADA), 1995-2013.....	235
8.2.1	Abstract.....	235
8.2.2	Introduction.....	236
8.2.3	Literature review.....	238
8.2.4	Methodology and research strategy.....	241

8.2.5	Case study and data	246
8.2.6	Results	254
8.2.7	Discussion.....	264
8.2.8	Conclusion.....	266
CONCLUSION GÉNÉRALE		268
RAPPEL DU CONTEXTE DE LA RECHERCHE.....		268
LA META-ANALYSE COMME STRATEGIE DE RECHERCHE.....		270
DES APPLICATIONS EMPIRIQUES		273
EN RESUME		276
RÉFÉRENCES BIBLIOGRAPHIQUES		278

LISTE DES TABLEAUX

Tableau 1.1: Proportion de publications par revue scientifique	25
Tableau 2.1: Les hypothèses du modèle de RLM (<i>Bourbonnais, 2003</i>)	46
Tableau 4.1: Articles considérant l'analyse en coupe transversale (omission du temps)...	119
Tableau 4.2: Articles considérant l'analyse spatiotemporelle	124
Table 5.1: Considered journals.....	153
Table 5.2: List of the considered variables in the meta-regression models.....	154
Table 5.3: Sample of included studies.....	157
Table 5.4: Descriptive statistics - 3 samples: Moran's I ; λ and ρ	159
Table 5.5: Estimation results	162
Table 6.1: Descriptive statistics.....	189
Table 6.2: Estimation results	194
Table 7.1: Temporal distribution of the transactions.....	220
Table 7.2: Spatial distribution of the transactions: proximity to the renovated street.....	222
Table 7.3: Estimation results	227
Table 8.1: Yearly distribution of sale and resale transactions (1995 to 2013 – 5612 pairs).....	250
Table 8.2: Temporal distribution of transaction	251
Table 8.3: Distance to the nearest metro station per project phase	253
Table 8.4: Estimation results	257
Table 8.5: Statistical tests on coefficients:	258

Table 8.6: Marginal effect of the effect of distance in dollars (\$)	262
--	-----

LISTE DES FIGURES

Figure 1.1: Champs de recherche	13
Figure 1.2: Trois perspective de l'espace géonomique (Perroux, 1950).....	14
Figure 1.3: Trois dimensions de l'espace géographique (Moran, 1963).....	18
Figure 1.4: Dimensions de l'espace géographique.....	23
Figure 1.5: Progression du nombre de publications	25
Figure 1.6 : Relations de dépendance spatiale et temporelle.....	31
Figure 1.7: Différents types d'organisation des observations	33
Figure 1.8: Observations en coupes transversales empilées (CTE)	36
Figure 2.1: Détermination du prix implicite pour l'attribut z1	42
Figure 2.2: Modélisation en contexte aspatial	48
Figure 2.3: Schéma conceptuel.....	54
Figure 3.1: Modélisation des processus spatiaux	93
Figure 3.2: Démarche de sélection du modèle économétrique.....	97
Figure 3.3: Variation expliquée par la méta-régression	111
Figure 5.1: Progression of research papers' publications (1984-2015).....	136
Figure 5.2: Spatial and temporal dependency relations.....	143
Figure 5.3: Frequency distribution of the datasets time ranges.....	161

Figure 5.4: Impact of the dataset temporal range on spatial statistics	164
Figure 6.1: Variation explained by the meta-regression model	184
Figure 6.2: Meta-analysis dataset building process	187
Figure 6.3: Spatial statistics' distributions	190
Figure 6.4: Comparison of the mean effects size for the Moran's I index	192
Figure 6.5: Comparison of the mean effects size for the Rho (ρ) coefficient	192
Figure 7.1: Spatial distribution of the repeated sales over the Old Quebec City	206
Figure 7.2: D'Auteuil street	209
Figure 7.3: D'Auteuil Street after restoration	211
Figure 8.1: City map and spatial distribution of the transactions	247
Figure 8.2: Project implementation timeline	248
Figure 8.3: Sale/Resale price evolution (1995-2013)	255
Figure 8.4: Proximity effect for each significant implementation phase	259
Figure 8.5: Total marginal effect expressed in dollars (CAD\$)	264

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

ACS	Autocorrélation spatiale
BLUE	Best linear unbiased estimator (<i>Meilleur estimateur sans biais</i>)
CBD	Centre-ville / quartier d'affaires (<i>Central business district</i>)
DID	Difference-in-differences estimator (<i>Estimateur de difference-en-differences</i>).
ET	Effet de taille (<i>Effect Size</i>)
GMM	Generalized method of moments (<i>Méthode des moments généralisés</i>)
GWR	Geographically weighted regression (<i>Regression géographiquement pondérée</i>)
LR	Likelihood ratio (<i>Ratio de vraisemblance</i>)
LSTAR	Local spatiotemporal autoregressive model. (<i>Modèle autorégressive spatial local</i>)
LWR	Régression locale pondérée (<i>Locally weighted regression</i>)
MA	Méta-analyse
MAUP	Modifiable areal unit problem (<i>Problème des aires modifiables</i>)
MCO	Moindres carrés ordinaires
ML	Maximum likelihood (<i>Maximum de vraisemblance</i>)
MPH	Modèle de prix hédoniques

MR	Méta-régression
PGD	Processus générateur de données
RLM	Régression linéaire multiple
SALE	Spatial autoregressive local estimation (<i>Estimation autorégressive spatiale locale</i>)
SAR	Spatial autoregressive model (<i>Modèle autorégressif spatial</i>)
SDID	Spatial difference-in-differences estimator (<i>Estimateur de difference-en-differences spatial</i>)
SEM	Spatial error model (<i>Modèle autorégressif d'erreur spatial</i>)
SIG	Système d'information géographique
SR	Science régionale
STHM	Spatiotemporal hedonic model (<i>Modèle de prix hédoniques spatiotemporel</i>)
TSLS/2SLS	Two-stage least squares (<i>Doubles moindres carrés</i>)

LISTE DES SYMBOLES

\$	dollars.
m	mètres.
km	kilomètres.
%	pourcentage.

INTRODUCTION

La notion de région est devenue un élément central dans le débat politique pour l'élaboration et l'application des politiques publiques (Hudson, 2007). À la fois objet et sujet des politiques publiques, la région constitue un centre d'intérêt dans la structure complexe des pouvoirs politiques et de l'espace de gouvernance. Néanmoins, depuis les travaux pionniers de Herbertson (1905), Fawcett (1919) ou De La Blache et al. (1926), celle-ci ne fait pas l'objet d'une définition claire et se heurte aux conceptions disciplinaires évolutives des territoires régionaux (Pike, 2007). Ainsi, pour Hudson (2007 : 1155), « Regions can be seen as products of complex condensations of social relationships, of varying density and variety, which combine contingently in specific time/space couplings to produce what are [...] unique regions with fuzzy and permeable boundaries ». De plus, les réflexions sur la notion d'espace, de lieux ou d'échelle tendent pour leur part à définir la région à partir de liens et relations à l'intérieur et à l'extérieur de territoires géographiques prédéfinis (Pike, 2007).

L'étude du phénomène régional et urbain représente donc à la fois un champ de recherche complexe, mais aussi éminemment important. La littérature peut être scindée en deux grands courants qui reposent sur des traditions méthodologiques distinctes. Le premier, les études régionales (ou *regional studies*), s'inscrit sur une perspective qualitative et des fondements disciplinaires éclectiques. Le second, la science régionale (ou *regional science*), est quant à lui associé à une approche empirique quantitative et d'inspiration principalement économique (McCann, 2007).

Au-delà des distinctions disciplinaires et méthodologiques, le regard porté sur le phénomène régional par ces approches s'oppose. Le premier répond à une approche normative et le second à une approche positive (*idem*). Pourtant, elles ne sont pas en concurrence. En fait, les deux approches posent plutôt des questions différentes: « these different types of approaches within urban and regional analysis should be seen as really asking different types of questions [...] not necessarily inconsistent with each other » (*idem*: 1210). L'auteur souligne que par nature les réflexions portées par les études régionales tendent à soulever des questions auxquelles la logique analytique de la science régionale tente de répondre par la suite.

Peu importe l'approche préconisée, la compréhension des dynamiques régionales et urbaines est une tâche complexe. Les dimensions temporelle et spatiale contribuent à contextualiser les phénomènes et problématiques de recherche. Ainsi, pour comprendre « pourquoi » un scientifique devrait répondre au « où » et au « quand » (Cressie et Wikle, 2011), la situation spatiale va implicitement de soi dans la perspective des territoires régionaux. L'engouement pour les méthodes d'analyse spatiale en témoigne. La considération simultanée de la situation temporelle, en revanche, peut faire défaut alors qu'elle est essentielle dans la définition des processus de causalité et donc dans la définition du « pourquoi ». En conséquence, une telle impasse remet en question la portée des analyses et pourrait de ce fait nuire à l'élaboration et l'évaluation de politiques publiques régionales.

La recherche vise à établir l'impact de l'omission de la dimension temporelle dans les analyses économétriques basées sur des microdonnées soulignant ainsi le caractère dynamique des phénomènes spatiaux (régionaux ou urbains). Une telle omission pourrait avoir pour effet d'invalider les résultats de recherche et, du même fait, de nuire aux analyses utiles aux politiques de développement régional. La thèse, de nature méthodologique, s'inscrit dans la tradition quantitative de la science régionale, mais tend à

entreprendre un lien entre les deux courants de l'analyse régionale et urbaine et la dimension temps, sous-tendant les dynamiques régionales et urbaines.

En effet, par l'application d'outils économétriques spatiaux appliqués à des données collectées à une échelle désagrégée, ce travail repose sur une conception réticulaire de l'espace ne correspondant pas à une définition rigide et homogène du territoire régional, autant dans l'espace que dans le temps. De plus, la recherche porte une réflexion sur l'usage de méthodes statistiques appliquées à des données de niveau individuel dans le but d'améliorer les méthodes d'évaluation de politiques publiques ainsi que la compréhension de phénomènes régionaux complexes préalablement à l'élaboration de telles politiques.

Les données immobilières sont utilisées comme exemple probant de données spatiotemporelles désagrégées non répétées. Néanmoins, les outils utilisés sont destinés à être appliqués à une grande variété de données ou de problématiques et à contribuer à l'amélioration des connaissances dans des thématiques de recherches variées (entreprises/commerces, criminalité, pauvreté, environnement, etc.).

L'utilisation de données immobilières n'est pas fortuite. L'analyse régionale et urbaine porte en effet une attention particulière à la localisation des activités. Depuis les travaux pionniers de von Thünen (en contexte agricole) et d'Alonso (en contexte urbain), l'espace est interprété comme une ressource dont la mobilisation est associée à une rente. Dans une perspective économique, les modèles basés sur la notion de rente foncière permettent notamment d'expliquer comment est organisée et structurée l'organisation de l'espace (Fujita, Krugman et Venables, 2001). Cette rente fluctue notamment en fonction d'externalités spatiales et affecte les valeurs foncières, tant pour les entreprises que pour le marché résidentiel.

Dans une perspective de développement régional, activité économique et lieu de résidence entretiennent des liens étroits jusqu'à la formation de zones urbaines denses :

« concentration of jobs lead to concentration of residences as people locate near their worksites. The result is a city » (Brueckner, 2001: 1). Les mécanismes du marché et le stock limité d'espaces disponibles contribuent à faire fluctuer grandement les coûts de la localisation en fonction de la distance, et affectent directement les budgets de ménages.

Ainsi, les valeurs immobilières captent les changements et l'évolution de leur contexte environnant. Elles subissent les transformations du contexte économique et social ou encore la mise en place de politiques publiques variées (infrastructures de transport, aménagement urbain, etc.). À titre d'exemple, pour un ménage le choix d'une localisation résidentielle est relatif à ses déplacements quotidiens. Les coûts de transport sont alors intégrés aux valeurs foncières (Kim et Lahr, 2014) qui peuvent être comprises comme un bon indicateur de mobilité (Burgess, 2008). En tant que parties intégrantes du milieu de vie, elles sont finalement un *outil* pertinent pour l'analyse de certaines dynamiques régionales et urbaines de même que pour la réflexion et l'implémentation de politiques publiques.

La thèse est divisée en neuf chapitres. Le premier vise à familiariser le lecteur avec le champ de la science régionale tout en l'introduisant aux principes de l'économétrie spatiale. Le second chapitre présente le cadre théorique de la recherche. La théorie hédonique ainsi que les bases de certains outils économétriques sont présentées. Le troisième chapitre propose pour sa part un aperçu plus méthodologique et technique des méthodes de modélisation spatiale et spatiotemporelle tout en mettant l'accent sur l'usage de microdonnées (données désagrégées). Par ailleurs, la démarche de recherche préconise l'utilisation de la méta-analyse qui y est également présentée. Une revue de la littérature est l'objet du quatrième chapitre et porte attention aux travaux empiriques misant sur une analyse hédonique strictement spatiale et spatiotemporelle. Les quatre chapitres suivants sont respectivement associés aux quatre articles de la thèse. Deux analyses exploratoires sont d'abord présentées (chapitres 5 et 6), suivies de deux applications empiriques (chapitres 7 et 8).

La première analyse exploratoire, une méta-analyse présentée au chapitre 5, est destinée à expliquer à partir de travaux empiriques publiés les variations dans certaines statistiques spatiales (*I de Moran* et coefficient autorégressif *rho* et *Lambda*) en fonction de la dimension temporelle. Ce travail repose sur une revue de la littérature exhaustive et sur une approche méta-analytique conventionnelle.

La seconde analyse exploratoire, présentée au chapitre 6, propose une méthodologie originale inspirée des approches méta-analytique et *Bootstrap*. Son objectif principal est de palier les limites de la première analyse (uniformisation des données, taille de l'échantillon, disponibilité des statistiques spatiales analysées, tester des approches de modélisation alternatives). Sur la base d'un échantillon de données empiriques (non simulées), des sous-échantillons aux caractéristiques prédéfinies sont aléatoirement créés afin d'estimer et calculer les statistiques spatiales d'intérêt (*I de Moran* et coefficient autorégressif *Rho*). Les variations dans ces statistiques sont ensuite expliquées par l'intermédiaire de méta-régressions.

Le chapitre 7 est une première application empirique concrète. Du point de vue de la thèse, ce travail vise à proposer une approche de modélisation basée sur l'estimation d'un modèle hédonique et d'un estimateur de différence-en-différences (DID). D'un point de vue empirique, l'étude de cas est destinée à évaluer une partie des impacts économiques d'une politique d'aménagement urbain dans le contexte d'un secteur historique (*le vieux Québec*).

Le chapitre 8 constitue la seconde application empirique. Dans le cadre de la thèse, cet article présente l'utilisation d'un modèle de prix hédoniques et d'un estimateur de différence-en-différences spatial (SDID). Ainsi, contrairement à l'article précédent les dynamiques spatiales endogènes sont intégrées au modèle. Par ailleurs, l'étude de cas vise également à évaluer l'impact économique de l'extension d'une infrastructure de transport (*Métro de Montréal vers Laval*). Contrairement à la majorité des travaux précédents, cette analyse fait l'hypothèse d'un effet non homogène dans le temps et l'espace de la nouvelle

infrastructure sur les valeurs foncières. Ainsi, l'effet est décomposé dans le temps (par période de construction) de même que dans l'espace (par nouvelle station de métro).

Finalement, un neuvième chapitre synthétise les résultats et propose une discussion.

CHAPITRE 1

SCIENCE REGIONALE ET ECONOMETRIE SPATIALE

1.1 ÉMERGENCE DE LA SCIENCE RÉGIONALE

1.1.1 Un vaste champ de recherches

La science régionale est une discipline plutôt jeune. Elle date de la fin des années 1940 et s'est largement inspirée de la macroéconomie en se consacrant sur l'analyse des phénomènes inscrits dans l'espace. La science régionale s'inspire largement de travaux pionniers tels que la théorie de la rente foncière (von Thünen, 1826 puis Alonso, 1964), la théorie de la localisation (Weber, 1909, Predöhl, 1928), la théorie des aires de marchés (Hotelling, 1929) ou encore la théorie des places centrales (Christaller, 1933, Lösch, 1944) et le principe des externalités d'Alfred Marshall (1890). Depuis son éclosion, la science régionale évolue au rythme de la compréhension du rôle de l'espace.

Dans la perspective de l'économie spatiale, l'espace constitue principalement un obstacle au transport et aux facteurs de production, et donc un facteur influençant les coûts de production et les fonctions de profits des entreprises. Les travaux portent sur la production agricole (Thünen, 1826), industrielle (Weber, 1909) ou sur la distribution spatiale des villes (Christaller, 1933). Au cours de ses différents développements et avec l'évolution de la compréhension du rôle de l'espace, la science régionale dépasse les limites disciplinaires économiques.

Dans les années 50, la science régionale apparaît dans un contexte de reconstruction et répond à « un besoin de société : contribuer, sur le plan scientifique, à la reconstruction économique de l'après-guerre » avec le soutien du gouvernement américain qui perçoit la science régionale comme un outil stratégique de localisation des entreprises et des ressources (Bailly, 2009: 2). Cette période représente une transition dans la conception de l'espace marquée par l'usage de la modélisation de l'organisation et de la croissance urbaine, puis par une période de recherche opérationnelle afin d'assister les gouvernements dans la distribution des services publics et la planification de l'utilisation du sol (Mills et MacKinnon, 1973; Funck, 1975 dans Bolton et Jensen, 1995).

L'affiliation de la science régionale à l'économie s'explique par la contribution majeure de Walter Isard (1919-2010) et de son intérêt pour l'analyse de la localisation. Isard a contribué à traduire les écrits des théoriciens allemands de la localisation, alors méconnus des économistes non germanophones (Boyce, 2004). Ce travail permit sans nul doute d'ancrer le futur champ de recherche de la science régionale sur les fondations préalablement développées par les théories de la localisation. Par la suite, Isard s'implique dans le développement d'un modèle régional d'*input-output* sur l'invitation de Leontief¹ et enseigne un cours d'économie axé sur les théories de la localisation et le développement régional à l'université Harvard. Il sera un ardent défenseur de l'analyse de la localisation et des problèmes régionaux auprès de l'*American Economic Association* (AEA), où il organise des sessions spéciales avec la contribution de sociologues et de démographes durant lesquelles est soulignée la nature interdisciplinaire de la science régionale (Boyce, 2004). À la suite de ces sessions de réflexions, la *Regional Science Association* (RSA) voit le jour en

¹ L'analyse *input-output* développée par Leontieff vise à associer théorie et pratique économique : «a method of analysis that takes advantage of the relatively stable pattern of the flow of goods and services among the elements of our economy to bring a much more detailed statistical picture of the system into the range of manipulation by economic theory» (Leontief, 1986)

1954 et prendra de l'envergure pour devenir la *Regional Science Association International* (RSAI) dont le programme est encore aujourd'hui décrit comme « an interdisciplinary mix of theory, methods and applications on both large and small regions, including urban and rural regions an interregional relationships » (*ibid.* : 34).

Suite à une rapide période d'expansion, le champ de recherche se retrouve confronté à une crise. La science régionale se heurte à une pensée globalisante d'une société ouverte au libéralisme. On passe alors d'une philosophie axée sur la région (penser régionalement) à une philosophie axée sur le global (penser globalement) (Bailly, 2009). La science régionale est donc remise en cause, ce qui mènera à la fermeture de départements entiers qui lui sont consacrés dans les universités. En 1994, le département de science régionale de l'université de Pennsylvanie, le premier à avoir vu le jour, ferme. Deux symptômes justifient l'identification d'une crise de la science régionale : i) le manque de pertinence sociale et ii) une perspective trop étroite. Selon Bailly et Coffey (1994), le manque de pertinence trouverait sa source dans un manque de connaissances théoriques influençant la qualité des travaux empiriques jusqu'à remettre en question la pertinence du financement de la science régionale. Les auteurs expliquent : « The past decade has seen the growth of both an increasing internal malaise due to a lack of emphasis on practical issues, and a corresponding lack of recognition in policy circles because the field has not sufficiently demonstrated that it can address real world problems » (Bailly et Coffey, 1994 : 4). Sous l'influence des méthodes quantitatives économiques, la période de l'après-guerre est marquée par une hausse des méthodes de modélisation en réponse à un manque de rigueur des sciences sociales dites « molles ». Néanmoins, la recherche d'une hausse constante de la rigueur se fait au détriment de la pertinence qui peut être en partie expliquée par des limites méthodologiques de l'époque.

En considérant la perspective de la géographie et de l'économie, Rees (1999) décrit les débats méthodologiques qui y règnent comme des opportunités pour le futur de la

science régionale de s'affirmer en misant sur un pluralisme méthodologique sans sombrer dans une anarchie épistémologique propre aux revendications postmodernes. L'auteur voit dans la tangente postmoderne un rejet des questions épistémologiques de la science et la notion de théorie elle-même : « The implication of postmodernism for regional science then would be to throw out all the theories and methods of spatial analysis as accumulated over the past half century » (Rees, 1999: 105). Du point de vue de Rosenau (1991), malgré leur rigidité, les cadres théoriques demeurent essentiels et leur disparition suppose l'homogénéisation des discours et, de ce fait, la disparition des débats. En conséquence, ce manque de pertinence identifié par Bailly et Coffey (1994) n'en serait que renforcé.

Ainsi, cette *introspection* épistémologique disciplinaire qui frappe pourrait être perçue comme une source d'opportunités pour la science régionale et ces débats méthodologiques se révéler une source d'inspiration. Néanmoins, la tâche demeure complexe : « no one seems to know how to halt the slide into the methodological anarchy of deconstruction without retreating to the rules of modernism » (Rees, 1999: 104).

On reproche également à la science régionale une perspective restreinte. À l'origine de la science régionale se trouve une volonté d'élargir les spectres disciplinaires classiques afin d'appréhender la région dans toute sa complexité. L'objectif de l'interdisciplinarité est accessible par un élargissement des perspectives en matière de disciplines, de méthodes ou encore d'objets d'analyse (Coffey, 1992). Malgré cela, elle conserve une dimension mathématique qui évacue plusieurs dimensions sociales et humaines et conduit la science régionale à une remise en question. L'individu doit être replacé dans les considérations des chercheurs comme Hägerstrand l'avait proposé lors d'une conférence présidentielle en 1973: « Regional science is about people, not just about location ». Isard (1960) a d'ailleurs lui-même défini la science régionale comme un champ de recherche multidisciplinaire qui examine la localisation de *toutes* les activités humaines associées à la société (économiques, sociales, politiques, etc.).

Il y a dans cette critique du champ de recherche un constat d'échec des scientifiques à répondre aux attentes actuelles des milieux (décideurs, praticiens, technologues, etc.). Bolton et Jensen (1995) identifient certaines de ces attentes envers les *regional scientists* parmi lesquelles se trouvent : i) le développement de cadres théoriques permettant de comprendre les phénomènes inscrits dans l'espace; ii) la recherche constante de tests et d'amélioration de ces théories et modèles afin de répondre aux besoins actuels des décideurs et iii) la considération de ces phénomènes tant au niveau régional qu'urbain.

1.2 ESPACE ET LOCALISATION

1.2.1 La notion d'espace dans l'analyse régionale

La région devient « objet d'étude spécifique dès lors que les préoccupations orientées vers l'action se font jour » dans différents pays et gouvernements (Aydalot, 1985: 11). L'analyse régionale voue une attention particulière à la notion d'espace dont la conception se transforme avec l'avancement des travaux et au rythme de l'évolution du contexte empirique. Malgré les travaux de précurseurs tels que von Thünen (1826) ou Ricardo (1824), la reconnaissance de l'importance de la dimension spatiale sur les processus économiques est encore récente et croissante (Duranton, 1997, Goodchild et al., 2000, Rey and Janikas, 2005).

Claval (2008) identifie trois périodes, marquées par deux transitions, dans l'évolution de la science régionale coïncidant avec des transformations de la notion de spatialité. La première période identifiée par l'auteur réfère aux travaux réalisés préalablement à l'éclosion de la science régionale. La première transition majeure est observée dans les années 50-60, suite à la Seconde Guerre mondiale. Cette période marque également l'apparition formelle de la science régionale. À partir des années 1970, une seconde

transition s'opère durant laquelle l'économie de l'offre, caractéristique du système productif fordiste, évolue vers une économie de la demande et supporte ainsi un besoin d'innovation constant. Ce contexte modifie le rapport à l'espace et contribuera à l'émergence de la notion de cluster. L'espace prend une nouvelle dimension et devient vecteur d'externalités : « l'espace n'est plus simplement traversé par des flux d'informations. Il est ponctué par des lieux où la mémoire s'accumule, où les connaissances se créent et où elles se perpétuent » (Claval, 2008: 175).

Ces changements de perspectives influencent les avancées théoriques. L'inclusion de l'espace dans l'analyse économique peut avoir deux conséquences principales : i) remettre en cause certaines théories économiques et ii) faire émerger de nouvelles problématiques. L'usage de microdonnées géolocalisées crée de nouveaux défis méthodologiques (Thisse, 1997) pouvant avoir des répercussions sur ces deux conséquences. L'utilisation croissante de la notion d'espace permet de décomposer l'analyse économique régionale en deux branches distinctes : i) les théories de la localisation, branche la plus ancienne et qui trouve ses bases dans la microéconomie; ii) les théories de la croissance régionale davantage influencées par le principe de la macroéconomie et axées sur les aspects spatiaux de la croissance régionale et la distribution territoriale des revenus (Capello, 2009b).

La notion d'espace n'est pas comprise de manière identique selon les problématiques, l'échelle d'analyse et le champ de recherche. Tandis que la géographie économique se définit comme un champ de recherche ciblant « la relation entre l'espace vécu et la vie économique » (Polèse et al., 2015: 1) incluant l'économie urbaine, la nouvelle économie géographique se définit comme un champ de recherche au croisement du commerce international, de l'économie urbaine et de la microéconomie spatiale destiné à « développer une théorie complète de la localisation des personnes et des activités, des flux entre ces localisations et des évolutions du paysage économique » (Duranton, 1997: 1). L'auteur ajoute que : « les sciences régionales, l'économie régionale et l'économie urbaine sont

autant de disciplines emboîtées dont le niveau de généralité diminue à mesure qu'augmente le niveau d'abstraction » (ibid.). Ces champs de recherche peuvent être compris comme un système de perspectives complémentaires dont les niveaux se distinguent en fonction des problématiques traitées et donc de l'échelle d'analyse privilégiée (figure 1.1).

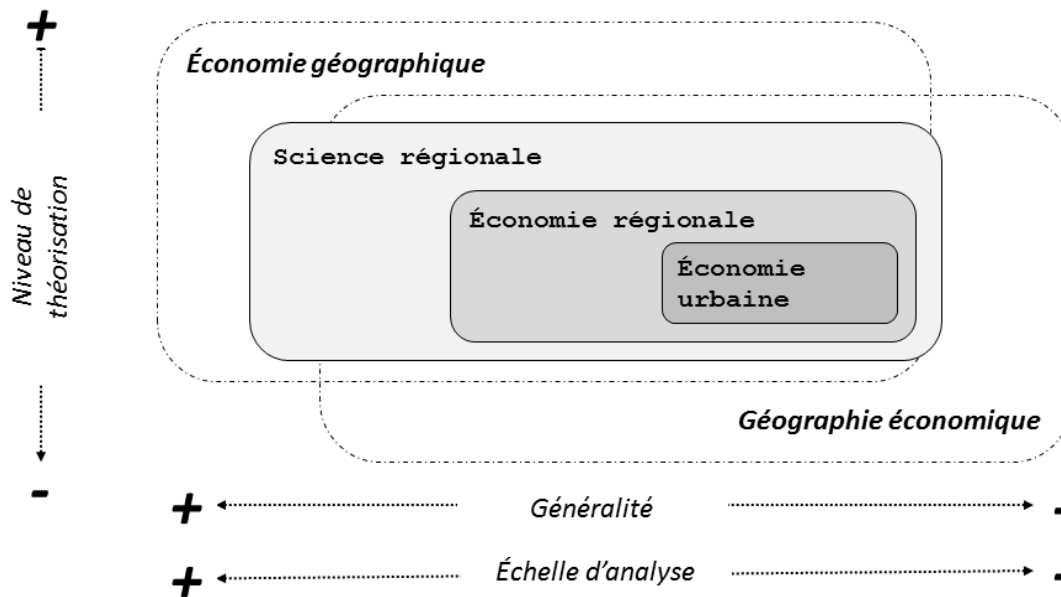


Figure 1.1: Champs de recherche

L'introduction de la notion d'espace géographique dans les théories ne constitue pas une transition achevée, mais plutôt un processus en cours qui affecte chacune de ces perspectives et champs de recherche. Également identifié comme une constituante de la science régionale (Bailly, 2009), le développement régional est un processus qui évolue au rythme des transformations de la notion d'espace (Aydalot, 1985, Claval, 2008) et il en va de même de manière générale avec l'économie géographique et ses composantes.

1.2.2 Définir l'espace géographique

Perroux (1950) décompose l'espace géonomique en trois catégories qui caractérisent les relations économiques (également présentés par Boudeville, 1968). L'espace peut alors être vu : 1) comme un plan, 2) comme un champ de forces, ou 3) comme un ensemble homogène.

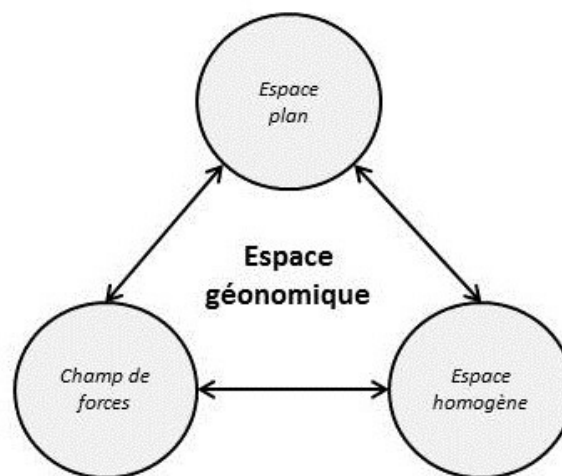


Figure 1.2: Trois perspectives de l'espace géonomique (Perroux, 1950)

Le premier type d'espace (espace plan) intègre l'ensemble des relations qui existent entre une firme, ses fournisseurs et son marché : « The economic space of the firm is thus in the first place a structure of relations defining the plan of this very unit and the plan for other units » (Perroux, 1950: 95). Cette conception pourrait être associée à l'interprétation vectorielle de Weber (1909) et au problème de localisation. L'auteur tente de déterminer une localisation optimale en fonction de points d'attraction qui exercent une force de localisation sur l'activité économique. Dans la décision de localisation, Weber intègre une

notion de profit influencée par les coûts de transport. Ainsi, pour Weber « maximiser le profit revient à minimiser le coût total de transport » (Tellier, 1993: 17). Weber identifie donc des points d'attraction sur un espace qui sont des sources d'intrants ou un marché.

Il est aussi possible d'associer à cette approche la théorie des aires de marché proposée par Hotelling (1929) selon laquelle le facteur de localisation principal est la proximité du marché en vue d'une maximisation du profit en tenant compte de la situation géographique des concurrents. Robinson (1931) souligne l'importance de la densité de la demande et de la surface du marché dans la décision de localisation. Le concept de demande spatiale se définit différemment de la demande classique (Tellier, 1993: 97) puisqu'il « exprime la demande d'un bien non pas en fonction du prix de ce bien, mais plutôt en fonction de la situation dans l'espace des consommateurs et du fournisseur ». Il est intéressant de constater que le modèle d'Hotelling intègre un univers qui n'est pas statique et dans lequel la localisation d'un agent peut influencer la localisation des autres.

Le second type d'espace (champ de forces) se rapporte directement à la théorie des pôles de croissance qui souligne le rôle de certains lieux centraux mais également d'une dynamique spatiale incarnée par des forces centripètes et centrifuges contribuant au déploiement ou à la contraction de l'activité (Perroux, 1950). Néanmoins, ces forces sont plutôt inscrites dans l'analyse du déséquilibre cumulatif proposée par Myrdal. Cette analyse met en évidence les effets contradictoires qui accompagnent une dynamique de développement (Courlet, 2001). La force centripète (que Courlet nomme effet de remous), est engendrée par un phénomène d'expansion d'une région produisant un régime de rendement croissant qui affecte les relations de concurrence et fait obstacle à la croissance des activités industrielles des régions moins développées. La force centrifuge produit l'effet inverse et devient bénéfique pour les régions voisines. L'activité économique d'une région est alors *projetée* vers sa périphérie, engendrant un processus de diffusion au profit des régions voisines.

Autrement dit, si la force centrifuge est plus importante que les effets de remous, de nouveaux centres de développement peuvent émerger. Dans un contexte urbain, la ville monocentrique verra des sous-centres émerger. Brueckner (2001) identifie trois forces fondamentales menant à l'expansion urbaine : i) la croissance de la population, ii) l'accroissement des revenus et iii) la diminution des coûts de transport. Dans ce contexte d'expansion, les effets d'attraction d'un centre urbain sont parfois imaginés comme une forme de magnétisme. Cependant, cette « théorie de l'aimant » idéalise un processus complexe et devrait inclure l'hypothèse d'effets de friction générés par l'espace, sans quoi la croissance urbaine et la densité de son centre seraient infinies (Duranton, 1999). Le modèle monocentrique correspond à un idéal typique des villes industrielles du 19e siècle : « researchers have tried in vain to adjust this standard monocentric model to the polycentric configuration of many modern cities » (Kloosterman and Musterd, 2001: 625).

C'est dans cet esprit que la théorie des pôles de croissance peut être associée à ces forces. Perroux (1964: 144) précise que « la croissance n'apparaît pas partout à la fois, elle se manifeste dans des points ou pôles de croissance, avec des intensités variables; elle se répand par divers canaux et avec des effets terminaux variables pour l'ensemble de l'économie ». Des économies externes résultent d'une économie d'agglomération décrite comme :

des gains de productivité attribuables à l'agglomération géographique des populations et des activités économiques. Comme la source de ces gains ne se trouve pas à l'intérieur de l'entreprise, dans sa technologie propre ou son mode de gestion, mais plutôt dans son environnement (sa localisation), on parle d'économies externes ou d'externalités positives. (Polèse and Shearmur, 2005: 55)

La théorie distingue deux types d'économies externes. La première est transmise horizontalement par l'intermédiaire du marché et touche l'ensemble des activités. Celle-ci réfère habituellement aux économies d'urbanisation. La seconde est transmise verticalement et propagée par l'activité qui l'a fait naître. Celle-ci réfère habituellement aux

économies de localisation. Ces deux types d'économies d'agglomération influencent ainsi les décisions de localisation, mais seraient également des facteurs endogènes influençant la croissance des agglomérations. Les relations et réseaux au sein d'une activité deviennent alors une source de croissance déséquilibrée selon Courlet (2001). Celui-ci ajoute que la notion de pôle de croissance est « liée à l'idée de relations fonctionnelles; elle a un contenu structurel et sectoriel avant d'avoir une signification spatiale » (2001: 17).

Le dernier type d'espace (ensemble homogène) se caractérise par des conditions de production, de concurrence, de vente, etc. :

It happens that various firms are placed in approximately the same conditions and set approximately the same price, for a clientele situated at the same physical distance. Alternatively, firms placed in very unequal conditions regarding cost, can offer the same price for clients situated at very different physical distances. These firms are in the same economic space (Perroux, 1950: 96).

Dans cette conception, l'espace est abordé de manière ponctiforme. Predöhl (1928), par exemple, compare la variation des prix des facteurs de production et les coûts de transport selon les lieux potentiels. Cette comparaison s'insère dans la perspective du principe d'utilité marginale et permet l'observation de points d'indifférence et de taux de substitution relatifs. Puisque l'espace est admis comme étant ponctiforme, un effet de substitution suppose une re-localisation de la firme (Ponsard, 1956). Le problème de localisation est lié à l'effet de substitution et devient donc un problème de prix (des facteurs de production et de coûts de transport) (Isard, 1949). Les limites de cette approche sont atteintes lorsque les sites ne sont plus uniquement considérés comme des « pourvoyeurs de ressources ou d'avantages comparatifs » (Zimmermann, 2008: 106).

Les trois espaces décrits par Perroux correspondent à trois approches distinctes de l'utilisation de la dimension spatiale dans les théories économiques. Les perspectives diverses de ces approches peuvent être attribuées à une compréhension différente de la

notion d'espace géographique. En effet, pour Moran (1963 : 24 dans Polèse et al., 2015), l'espace géographique (figure 1.3) est une notion complexe qui peut être comprise selon trois dimensions : les lieux, la surface et la distance. Ces trois dimensions et leurs interactions ont progressivement intégré la théorie économique et contribué au développement des théories de la localisation dans les champs de l'économie urbaine et de la science régionale (Fujita et al., 2001).² Considérés individuellement, les lieux ne permettent pas d'intégrer complètement l'espace aux développements théoriques. L'espace est d'abord abordé de manière ponctiforme et la mise en relation des lieux par l'entremise de la distance est un premier pas vers l'élaboration des théories de la localisation.

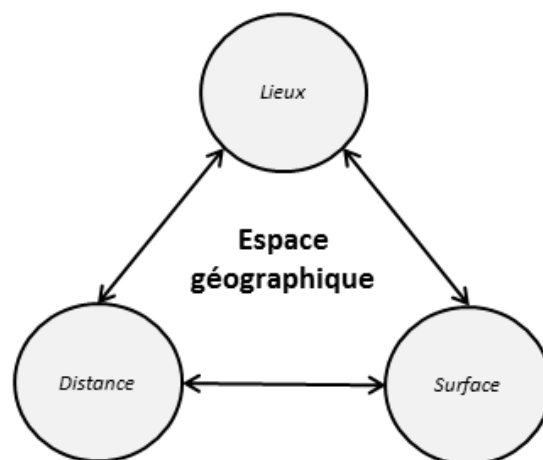


Figure 1.3: Trois dimensions de l'espace géographique (Moran, 1963)

² Fujita et Krugman ont d'ailleurs contribué à jumeler ces deux courants d'analyses (science régionale et économie urbaine) en tentant de définir la source des externalités d'agglomération et d'urbanisation (Claval, 2008).

La distance est certainement le premier élément utilisé afin d'inclure dans l'analyse économique l'espace géographique auparavant ignoré. Isard a par exemple souligné l'intérêt des travaux néo-classiques pour la dimension temporelle jugée cruciale au détriment de la dimension spatiale. Marshall prétend d'ailleurs : « The difficulties of the problem depend chiefly on variations in the area of space, and the period of time over which the market in question extends; the influence of time being more fundamental than that of space » (Marshall, 1920 dans Capello, 2009 : 34).

La distance est une forme de friction qui peut être économiquement quantifiée, notamment par le biais des coûts de transport (Polèse et al., 2015). Ainsi, l'inclusion de la distance à la compréhension des phénomènes économiques symbolise les débuts de l'insertion de la dimension spatiale dans les théories économiques. Isard critique Hicks et souligne son rôle dans la considération de l'espace à la théorie économique : « [Hicks] confines himself to a wonderland of no dimensions [...] in this sense the factor of space is repudiated, everything within the economy is in effect compressed to a point, and all spatial resistance disappears » (Isard, 1949: 477).

L'utilisation complémentaire des lieux et de la distance contribue à l'affirmation de la science régionale et des théories de la localisation : « Distance is the fundamental concept, the distinguishing element of our science » (Beckmann, 1976: 1). Duranton (1999) décrit la distance comme un facteur régulateur de la croissance urbaine. Sans les effets de celle-ci, les villes seraient vouées à croître indéfiniment, selon l'auteur qui nomme ce principe *tyranny of distance*. Plus récemment, les progrès technologiques ont encouragé une remise en question de l'importance de la distance comme déterminant de la localisation. Néanmoins, le principal effet des nouvelles technologies est une baisse des coûts de transport de l'information, pour le reste : « distance is still very much alive » (Polèse and Shearmur, 2004: 431). Pourtant, si la contribution de la distance n'a pas disparu, elle a connu des transformations au fil du temps.

À ce chapitre, l'interprétation vectorielle du problème de la localisation par Weber (1909) contribue à en faire un pionnier de l'analyse économique spatiale. Celui-ci tente de définir une localisation optimale en fonction de points d'attraction qui exercent une force, dite force de localisation, sur l'activité économique, lui permettant de lier coûts de transports et profits dans la décision de localisation.

Les dimensions de la distance et de la surface peuvent être utiles dans la compréhension de l'organisation ou de la structure spatiale des phénomènes économiques. À titre d'exemple, la théorie des places centrales (Christaller, 1933, Lösch, 1944) se résume à un ensemble de propositions permettant d'expliquer certaines préférences de localisation de l'activité économique dans une structure hiérarchisée et qui sont répétées à travers l'espace (Mulligan et al., 2012). La théorie des aires de marché de Hotelling (1929) considère également l'espace géographique dans cette perspective. La distance et la surface ont également inspiré la théorie de la rente foncière (Alonso, 1964, Thünen, 1826). Les progrès dans le transport ont certainement réduit l'effet de friction causé par la distance, mais au profit d'une forme de régulation de l'espace par le sol, un principe que Duranton (1999) nomme *tyranny of land*.

À mesure que les théories ont évolué, la notion de lieu s'est vue transformée, affectant les différentes approches théoriques avec l'apparition des notions de concentration et de dispersion. Les lieux auparavant considérés de manière ponctiforme ne sont plus uniquement considérés comme des éléments intégrés et homogènes.

Ullman (1958) observe qu'en pratique la concentration dans l'espace ne se résume pas à un point mais plutôt à une aire de concentration de points, contrairement à ce que Haig a théoriquement postulé : « the concentration is at points -cities- in an area » (Ullman, 1958: 197). Ce changement de perspective introduit une mutation du rôle de l'espace dans la théorie économique. Du point de vue de la théorie de l'échange international, l'espace constitue un facteur de production et de mobilité des biens d'un *point* à un autre. Suite à

l'évolution de la notion de lieu, les lieux tendent à être interprétés de manière plus contextuelle. Ils deviennent des endroits chargés de sens et structurés par des interactions créatrices « de ressources productives dans une dynamique fondée sur la combinaison de logiques endogènes et exogènes de développement » (Zimmermann, 2008: 115). Ils sont porteurs d'externalités autour desquelles s'organise l'activité humaine de manière dynamique. L'espace géographique est alors hétérogène et se distingue par l'émergence d'effets de concentration ou de débordement qui le transforment.

Dans cette perspective, la définition de l'espace est principalement orientée vers son contenu. D'ailleurs, la géographie économique caractérise davantage l'espace par la propriété des lieux qui sont de nature plus complexes (Claval, 2008). À ce stade, la notion de distance s'est transformée pour souligner le rôle régulateur des interactions personnelles directes, un principe que Duranton (1999) nomme *tyranny of proximity*.

Dans cette perspective, l'utilisation de concepts tels que ceux d'économies externes et d'agglomération, ou encore d'effets de débordements (*spillovers*) prennent tout leur sens. Ils contribuent à définir les lieux de manière continue (par opposition à une approche ponctuelle). S'en est suivi l'apparition de vastes champs de recherche propres à l'économie urbaine et aux milieux innovateurs / pôles de croissance / clusters.

Ces changements indiquent une transition importante dans le niveau d'analyse et l'approche des problématiques régionales. À titre d'exemple, le contexte urbain constitue un autre niveau d'analyse pour les modèles de localisation (Polèse, 1994). Contrairement à un espace « macro », caractérisé par une variation de ses attributs (ressources, coût de la main-d'œuvre, etc.), sous le prisme économique la ville représente un espace plus intégré, précise Polèse. Capello et Nijkamp (2004) ajoutent que, par nature, les villes sont des systèmes complexes où les aspects sociaux et économiques sont en interaction. D'ailleurs, certains travaux sur la croissance régionale tendent à montrer que les effets de

débordements et la localisation physique importent davantage que les facteurs macro-économiques traditionnels (Rey and Janikas, 2005).

Une échelle d'analyse plus fine aura tendance à octroyer aux analyses un niveau de particularité grandissant. À partir de 1980, les notions marshalliennes de district industriel, d'atmosphère industrielle et d'externalités sont réintroduites par les travaux italiens de Bagnasco, Garofoli, Becattini et Camagni, puis par le GREMI. Les notions de milieux innovateurs et de grappes d'activités (clusters) sont mises en évidence dans les travaux de de Scott et Storper, Piore et Sabel, Porter, Benko, Lipietz ou encore Courlet. L'espace est alors caractérisé par l'apport de la proximité des acteurs et prend une nouvelle consistance en devenant vecteur d'externalités. Parmi ces externalités, les effets positifs de débordement de la connaissance sur l'innovation et la croissance stimulent de manière significative les productions scientifiques : « they are at the heart of modern cluster theories [...] and the regional innovation system approach » (Maier and Trippel, 2009: 59).

L'espace compris comme un contenu se rapproche de la notion de territoire. Celle-ci répond à une prise de conscience de la contribution potentielle du local et du régional à générer un dynamisme favorable au processus de développement, une sorte d'endogénéité des variables qui étaient, jusque-là, surtout considérées comme exogènes (Massicotte, 2008). Les lieux prennent alors une nouvelle consistance et deviennent vecteurs d'externalités où des connaissances sont générées et se perpétuent (Claval, 2008). Le territoire confère à l'espace une certaine *épaisseur* en intégrant, selon Moine (2006: 117), tout « ce qui nous entoure dans une acceptation très globale » et que nous pouvons concevoir de manière intégrée. Massicotte (2008: XV) ajoute que le « concept de territoire [...] permettait en effet de se donner comme objet des lieux de dimension et de localisation variables [...] tout en facilitant la mise en évidence des interactions entre les territoires par imbrication, proximité ou relations diverses ».

De manière générale, l'élaboration des théories propres à l'analyse régionale requiert la mise en relation de ces trois dimensions (lieux, distance et surface) selon un équilibre variable. Malgré l'évolution de ces théories et la conception changeante de la notion d'espace, les dimensions de l'espace géographique demeurent au fondement de celle-ci³ (figure 1.4). La conception des lieux comme contenu et contenant constituent donc des approches opposées associées à des échelles d'analyse distinctes. Au-delà des théories de la localisation, beaucoup d'efforts restent à faire pour introduire l'espace à l'analyse régionale. La tendance axée sur l'usage de données désagrégées pour une analyse des dynamiques régionales et territoriales est l'un de ces défis.

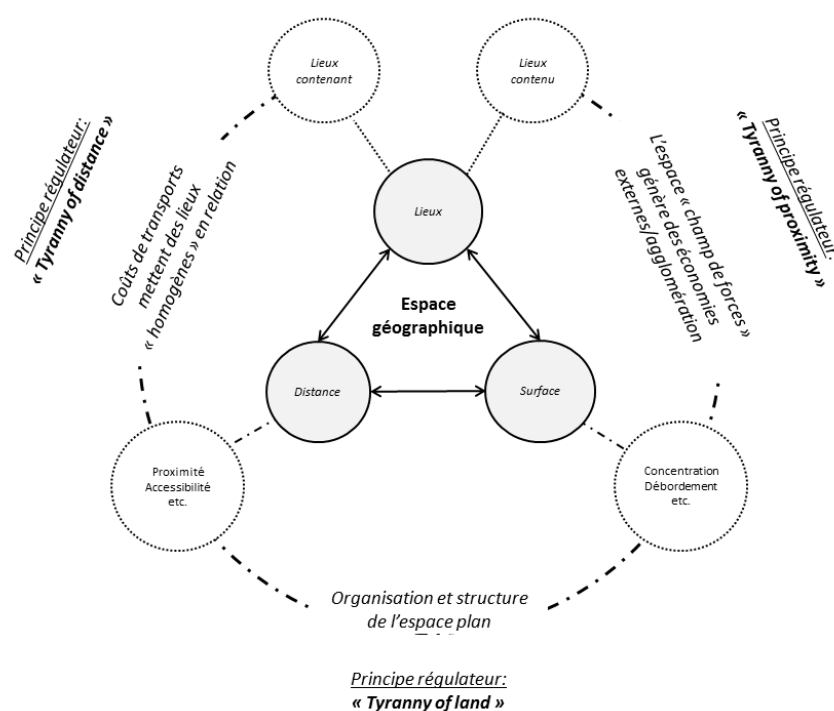


Figure 1.4: Dimensions de l'espace géographique

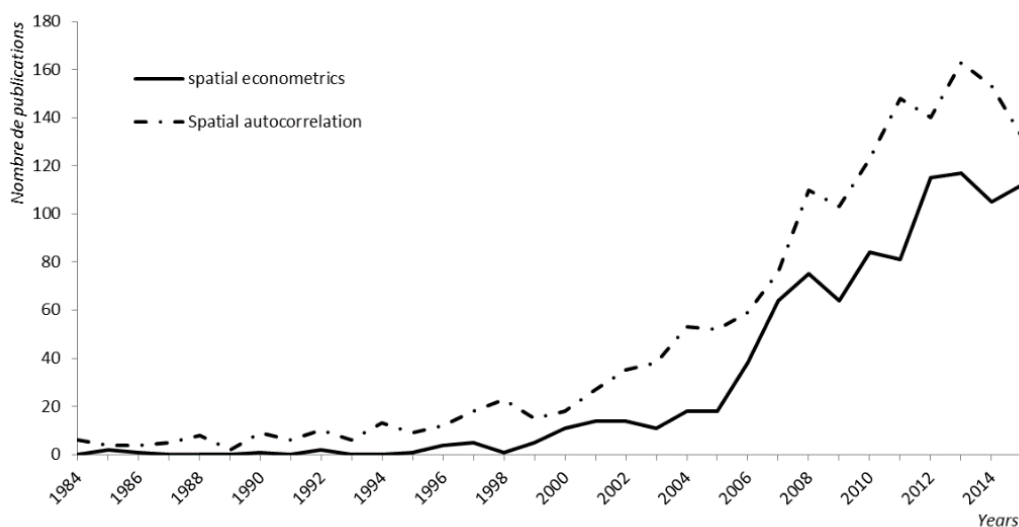
³ En aucun cas les différentes approches proposées ne sont mutuellement exclusives ni ne possèdent de délimitations marquées.

1.3 CONTRIBUTION DE L'ÉCONOMÉTRIE SPATIALE

1.3.1 Développement de l'économétrie spatiale

Qu'ils soient de nature économique ou sociale, les phénomènes régionaux sont complexes et leur analyse implique de recourir à des méthodologies de recherche variées et parfois innovantes. Certaines de ces démarches, associées à l'économétrie spatiale par exemple, sont relativement récentes et en constant développement. De manière générale, l'économétrie vise à appliquer diverses méthodes statistiques à l'économie afin de tester les théories et d'évaluer la contribution de paramètres dans les relations économiques (Dormont, 2007).

L'économétrie spatiale vise à modéliser des phénomènes situés dans l'espace et sa logique repose sur la première loi de la géographie selon laquelle « Everything is related to everything else, but closer things more so » (Tobler, 1979: 236). Bien au-delà de l'économie, les méthodologies qui y sont développées s'avèrent utiles dans un grand nombre de champs de recherche ou d'applications scientifiques (Arbia, 2011) et se prêtent particulièrement à l'analyse des phénomènes territoriaux notamment par l'usage de microdonnées (Dubé and Devaux, 2014). Avec la disponibilité croissante de données géolocalisées, le nombre de productions scientifiques associées à l'analyse spatiale subit une importante progression depuis la fin des années 80 jusqu'à devenir un champ incontournable. Le nombre de publications scientifiques manipulant ces méthodes est croissant (figure 1.5), malgré les critiques qui sont formulées face à cette nouvelle branche de l'économétrie (Gibbons and Overman, 2012, McMillen, 2012).



Note: La figure présente la croissance du nombre de publications associées aux mots clefs "spatial econometrics" et "spatial autocorrelation". Une importante hausse est observée à fin des années 90 et souligne l'intérêt croissant pour ces méthodes d'analyse.

Figure 1.5: Progression du nombre de publications

Le mot clef « autocorrélation spatiale » (spatial autocorrelation) réfère à une grande variété d'applications comme l'illustrent les principales revues ciblées (tableau 1.1). En précisant la recherche au mot clef « économétrie spatiale » (spatial econometrics), il appert que les trois premières revues en termes de nombre d'articles publiés se revendiquent du champ de la science régionale.

Tableau 1.1: Proportion de publications par revue scientifique

Mot Clef	Rang	Revue	%
Spatial Econometrics (1978-2012; N=538)	1	Regional Science and Urban Economics*	7,1
	2	Papers in Regional Science*	5,9
	3	Journal of Regional Science*	4,1
	4	Journal of Geographical Systems	3,3
	5	American Journal of Agricultural Economics	2,8

*Revue directement associée à la science régionale.

Plusieurs grandes périodes peuvent être identifiées dans le développement des méthodes d'analyse spatiale. Anselin (2010) identifie principalement trois phases dans le développement de l'économétrie spatiale : i) la période de mise en place des conditions préalables à son éclosion; ii) la période d'envol de la discipline, iii) la période de maturité de la discipline.

La reconnaissance et la conceptualisation du phénomène de dépendance spatiale sont parfois considérées comme formant une période supplémentaire antérieure à celles proposées par Anselin (2010). En effet, plusieurs auteurs ont suspecté la présence d'une forme de dépendance entre des données géographiquement proches. Student (1914) est généralement mentionné comme étant le premier à en faire mention dans ses travaux (Griffith, 2012, Le Gallo, 2002). Yule (1926) pose une hypothèse similaire en mentionnant un biais de mesure de la corrélation entre les variables en raison de l'autocorrélation spatiale. Stephan (1934) illustre le phénomène ainsi: « Data of geographic units are tied together, like bunches of grapes, not separated, like balls in an urn » (Stephan, 1934 : 165 dans Griffith, 2012). Finalement, Fisher (1935) constate également que les observations spatialement proches ont tendance à se ressembler.

La première période décrite par Anselin (2010) est marquée par la révolution quantitative en géographie et par le besoin d'incorporer les effets spatiaux aux modélisations dans le champ de la science régionale et de l'analyse économique régionale et urbaine. Cette période est marquée par le développement de mesures de cet effet de dépendance spatiale (autocorrélation spatiale) ainsi qu'au contrôle de celle-ci dans les modèles statistiques.

L'indice I de Moran et le coefficient c de Geary sont les deux principales statistiques élaborées pour détecter la présence de patron spatial dans un jeu de données. Pour l'heure, la statistique de Moran s'est révélée la plus puissante statistiquement (Griffith, 2012). Cliff

et Ord (1973) réaliseront un état des savoirs en statistiques spatiales qui servira d'impulsion au développement de nouvelles méthodes d'analyse. Ils ont développé une théorie de la distribution de ces deux statistiques pour les résidus de la régression linéaire associée à un modèle de probabilité de distribution normale et permettent de cristalliser l'ère des tests d'hypothèses pour l'autocorrélation spatiale (Griffith, 2012: 4). Suite à ces avancées, Besag (1974, 1975) serait le premier à utiliser la terminologie de *modèles autorégressifs*, tandis que Cliff et Ord contribuent à la spécification et à l'estimation de ces modèles auxquels le coefficient de Moran est désormais associé.

Cliff et Ord (1968) inventent le terme d'*autocorrélation spatiale* qui auparavant était identifié par les termes de *dépendance spatiale*, *interactions spatiales*, ou encore *interdépendance spatiale* (Getis, 2008). Pourtant, Anselin et Rey (1991) invitent à distinguer l'*autocorrélation spatiale* de la *dépendance spatiale*; la première se rapporte à une notion statistique, tandis que la seconde réfère au processus et aux structures spatiales qui organisent les données.

Les travaux de Ord (1975), Paelinck et Klassen (1979), Anselin (1980) permettent de préciser le premier cadre d'analyse de Cliff et Ord (Le Gallo, 2002). D'ailleurs, Paelinck et Klassen (1979) sont à l'origine du terme *économétrie spatiale* (*spatial econometrics*) qui fait la transition avec la statistique spatiale de Cliff et Ord (Arbia, 2011, Griffith, 2012) en ciblant les données économiques. Suite à cette nouvelle impulsion, des ouvrages spécifiques sur le cheminement des méthodes en économétrie spatiale sont publiés. En effet, le phénomène d'autocorrélation spatiale est en premier lieu abordé par les travaux considérant les externalités spatiales, mais est ensuite étendu à tous les domaines de recherche misant sur des données localisées⁴.

⁴ Anselin et Griffith (1988) fournissent une revue de la littérature de travaux provenant de disciplines variées manipulant des méthodes d'analyse spatiale.

Durant la période d'envol de la discipline, l'économétrie spatiale gagne en rigueur. Les changements les plus importants sont constatés dans l'analyse empirique dans les champs de l'économie régionale, urbaine et du développement, de l'économie publique, de l'analyse foncière et de l'économie du travail (Anselin, 2010). La période est marquée par des avancées techniques d'estimation (ex. : méthode des moments généralisés) qui deviennent standards.

Devant le renouveau de l'intérêt pour l'espace et les interactions spatiales dans les théories de sciences sociales, Anselin et Florax (1995) rédigeront l'ouvrage *New directions in spatial econometrics* qui constitue un état de la littérature complet sur les développements dans les méthodes économétriques intégrant l'espace (spécification des modèles, tests et estimateurs des modèles de régression, effets spatiaux sur les modèles statistiques, etc.). En réponse à une littérature grandissante sur le sujet, Anselin, Florax et Rey (2004) renouvellent cet exercice en publiant *Advances in spatial econometrics: Methodology, tools and applications* afin de faire une mise au point des développements depuis 1995:

In the few years since *New directions* appeared, [...] a tremendous growth in the number of publications in which spatial econometric techniques are applied, not only within regional science and economic geography, but also increasingly in the leading journals of economics, sociology and political science (Anselin et al., 2004: 1).

Finalement, Anselin (1988), LeSage et Pace (2009b) ou encore Elhorst (2014b, 2014a) publient des ouvrages méthodologiques de référence qui contribuent à stabiliser cette jeune discipline qu'est l'économétrie spatiale. Anselin (1988) y souligne la nécessité de tenir compte de l'autocorrélation spatiale dans l'utilisation de modèles et données géoréférencées (Getis, 2007). Néanmoins, les travaux empiriques n'ont pas rapidement intégré les avancées méthodologiques associées au traitement de données spatiales: « even though the methodological results achieved in the fields of spatial statistics and spatial

econometrics have been substantial, the dissemination from the research community to the applied world has been virtually nonexistent » (Anselin and Griffith, 1988: 14). Plus récemment, Rey et Janikas (2005: 156) font le même constat: « Methods developed in the mainstream of social science disciplines have been applied with little regard for the potential challenges posed by spatially referenced data ».

Pourtant, la discipline atteint une période de maturité grâce aux progrès informatiques et au développement de logiciels (Anselin, 2010). La discipline n'a pas fini d'évoluer au rythme de l'amélioration des méthodes d'économétrie spatiale, jusqu'à considérer de manière croissante la dimension temporelle. Arbia (2011) réalise une revue de la littérature des travaux théoriques et appliqués dans les cinq premières années d'existence de la *Spatial Econometrics Association* (2007 à 2011). Il identifie clairement dans les développements théoriques une transition graduelle de l'analyse en coupe transversale des données spatiales vers une modélisation de données en panel (démarche spatiotemporelle). En revanche, peu d'attention semble portée sur les données collectées de manière non répétées dans le temps et l'espace.

1.3.2 Phénomène de dépendance spatiale

Il existe deux formes de dépendance spatiale. La première met en relation les résidus du modèle de régression, tandis que la seconde souligne les liens entre des variables d'intérêt (Anselin, 2002). La notion d'autocorrélation caractérise la mesure de cet effet de dépendance. Tandis que la corrélation statistique *classique* désigne la relation entre deux variables, l'autocorrélation réfère à la corrélation entre les valeurs des différentes observations pour une même variable. L'autocorrélation spatiale d'une variable n'est pas un problème en soi (Legendre, 1993). Le problème vient plutôt lorsque les résidus d'un modèle de régression sont autocorrélés. Ce comportement viole alors l'hypothèse

d'indépendance et invalide alors l'interprétation des résultats obtenus par le modèle de régression : « it impairs our ability to perform standard statistical tests of hypotheses » (Legendre, 1993). Anselin et Griffith (1988: 16) précisent : « The regression coefficients will remain unbiased using OLS, but significance tests and measures of fit may be misleading ». En conséquence, la précision des estimations, et donc des conclusions qui peuvent en être tirées, est affectée (Elhorst, 2001).

Une seconde partie de la problématique réside dans la manière adéquate de tenir compte des effets de dépendance qui structurent les valeurs mesurées, particulièrement dans les bases de données où espace et temps se côtoient. L'autocorrélation spatiale constitue une forme de dépendance spatiale entre les observations en fonction de leur localisation géographique à un moment particulier dans le temps. En d'autres termes, ce phénomène se décrit comme la coïncidence entre les valeurs observées selon la localisation (Anselin and Bera, 1998, Chasco and López, 2008, LeSage and Pace, 2009b). La dépendance spatiale peut être positive (valeurs d'observations spatialement proches similaires) ou négative (valeurs d'observations spatialement proches opposées).

Les dépendances spatiale et temporelle sont de natures différentes (figure 1.6). Dans le cas d'une série de données chronologiques, les interactions entre les observations sont unidirectionnelles dans la mesure où les observations actuelles peuvent influencer celles à venir, alors que l'inverse est impossible (figure 1.6a). Dans le cadre de données spatiales, non seulement les interrelations sont réciproques, mais elles respectent également une logique multidirectionnelle (figure 1.6b) qui complexifie l'analyse (Le Gallo, 2002).

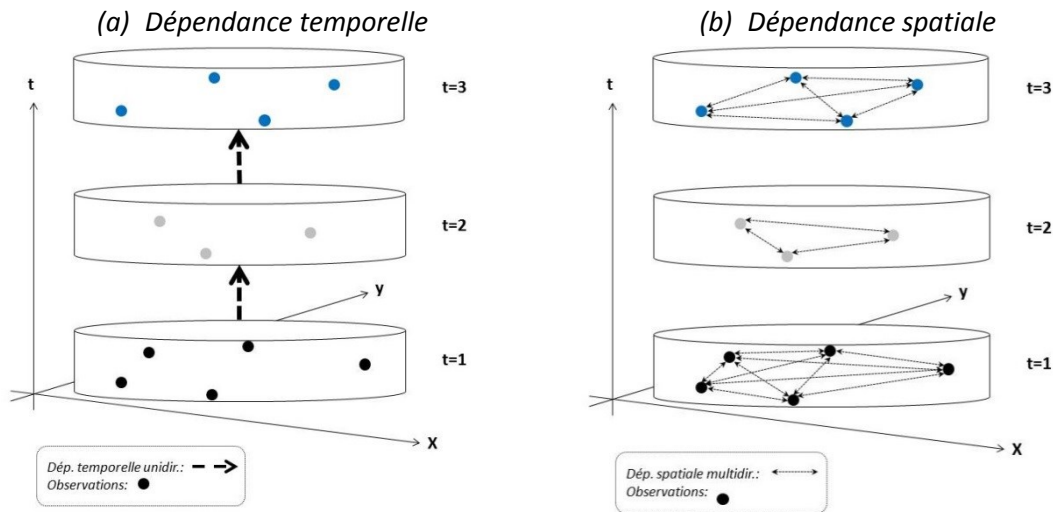


Figure 1.6 : Relations de dépendance spatiale et temporelle

Néanmoins, il faut distinguer ces deux types d'autocorrélation puisque: 1) la nature multidirectionnelle de l'espace implique de considérer des relations simultanées provenant de toutes les directions; 2) contrairement à l'unité de mesure temporelle, il existe une grande variété de mesures utilisables dans le contexte spatial permettant d'interpréter une forme de proximité (distance, contiguïté, liens, etc.). Finalement, l'autocorrélation spatiale constitue un concept incontournable dans les études utilisant les statistiques dans un contexte spatialisé « No other concept in empirical spatial research is as central to model building as is spatial autocorrelation » (Getis, 2008: 299). Pourtant, les méthodes développées dans le courant des sciences sociales en général sont appliquées avec peu de considérations pour les défis posés par celle-ci ou pour ses effets dans le temps (Ye and Rey, 2013). Getis (2007: 493) énumère les éléments justifiant la considération de l'autocorrélation spatiale dans l'analyse de données géoréférencées :

- fournir des tests pour la mauvaise spécification des modèles;
- déterminer la force de l'effet spatial sur les variables du modèle;
- permettre des tests sur les hypothèses de stationnarité et d'hétérogénéité spatiale;
- trouver les relations de dépendances potentielles;
- déterminer le rôle de l'éloignement ou des interactions spatiales sur les modèles spatiaux autorégressifs;
- aider à définir l'influence de l'organisation des unités géographiques sur les variables étudiées;
- permettre de définir la force d'association entre la réalisation de variables d'unités géographiques différentes;
- donner les moyens de réaliser des tests d'hypothèses sur les relations spatiales;
- permettre la pondération de l'importance des effets temporels;
- centrer l'analyse sur les interactions géographiques locales;
- aider à l'étude des variables aberrantes.

1.4 PROBLÉMATIQUE

1.4.1 Échelle d'analyse et structure des données

Isard (1960) a lui-même défini la science régionale comme un champ de recherche interdisciplinaire qui examine la localisation de *toutes* les activités humaines associées à la société (économie, sociale, politique, etc.). L'utilisation des méthodes économétriques spatiales appliquées aux microdonnées (individus, firmes, valeurs résidentielles, etc.) permet de renouveler la pertinence de méthodes mathématiques en appréhendant les processus et dynamiques qui les structurent. Dans cette perspective, la recherche proposée vise à dépasser certaines pratiques méthodologiques devenues problématiques. Plus précisément, il est ici question du traitement économétrique de données spatiotemporelles collectées de manière non répétée et analysées en coupes transversales empilées dans le temps.

Identifié comme une constituante de la science régionale (Bailly, 2009), le développement régional est un processus qui évolue au rythme des transformations de la notion d'espace (Aydalot, 1985, Claval, 2008). La notion de territoire se solidifie dans le prolongement de ces transformations. Néanmoins, le cumul des dimensions spatiale et temporelle complexifie la structure des processus structurant les observations. De manière générale, il existe deux grands types de structures des données. Le premier, celui des données en coupes transversales, ne considère que des données collectées à une seule période. Le second, celui des données de panel, intègre la dimension temporelle et permet un suivi chronologique des observations. La figure 1.7 présente les différentes structures possibles des observations en fonction du type de données utilisées (voir Dubé and Legros, 2014a).

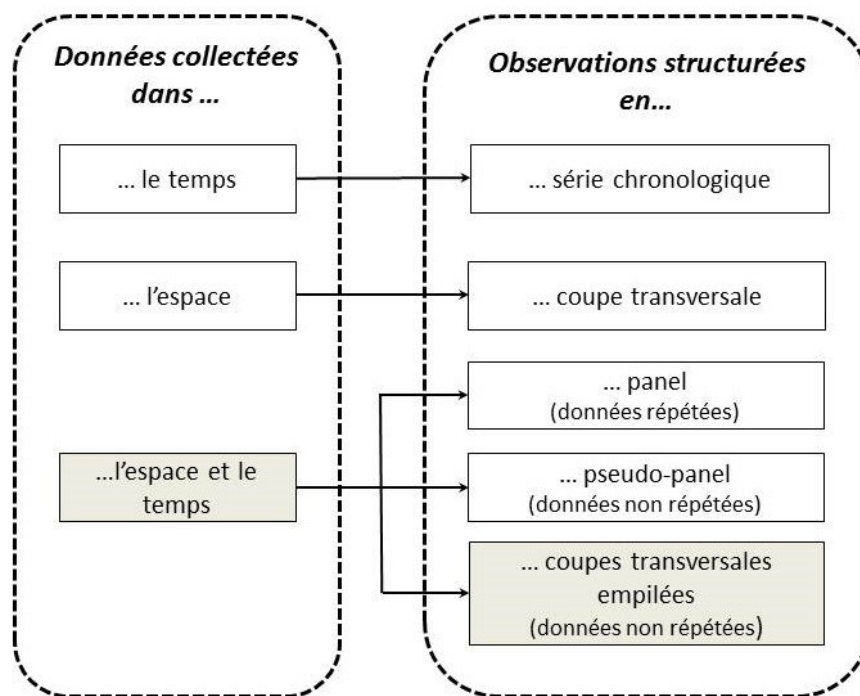


Figure 1.7: Différents types d'organisation des observations

L'approche en coupe transversale ne permet pas d'exploiter toute l'information comprise dans les données et permettant d'identifier la nature précise des mécanismes de causalité sous-jacents. Il est alors impossible, par exemple, d'identifier si un cluster provient d'un processus de diffusion ou d'une hétérogénéité spatiale ou structurelle (Anselin, 2002: 250). Les observations sont alors généralement traitées sous la forme de panel, c'est-à-dire localisées dans l'espace et répétées dans le temps. Anselin et Griffith (1988) précisent que si l'attention de l'économétrie a longtemps été portée sur les séries chronologiques, les possibles relations de dépendance entre les coupes transversales ont longtemps été ignorées. La répétition des observations dans le temps permet de définir les processus de dépendance spatiotemporels qui illustrent les relations qu'elles entretiennent.

En théorie, l'application des méthodes d'économétrie spatiale aux microdonnées permet donc de répondre au besoin de flexibilité d'échelle et d'intégrer les dynamiques spatiotemporelles telles que celles qui caractérisent les phénomènes régionaux. Pourtant, certaines données ne peuvent être collectées en continu. Les microdonnées peuvent être confrontées à cette difficulté (individus, firmes, logements, etc.). Il est alors proposé de les considérer sous la forme d'un pseudo-panel (ex. : Huang et al., 2006), c'est-à-dire en agrégeant les observations collectées de manière non répétée par région qui, elles, sont géographiquement stables dans le temps. En d'autres termes, tout en étant similaires le pseudo-panel se distingue de l'analyse en panel en intégrant des groupes d'observations non répétées selon des caractéristiques semblables et invariantes dans le temps ou l'espace : « In pseudo-panel analyses, individuals are grouped according to criteria that do not change from one survey to another » (Gardes et al., 2005: 242). Ainsi, contrairement au panel, la composition des aires spatiales considérées change d'une période à l'autre. Cette démarche permet de postuler des processus spatiotemporels, mais engendre des pertes d'informations causées par l'agrégation et confronte le chercheur au problème des aires modifiables (*Modifiable Areal Unit Problem*, MAUP - voir Openshaw, 1983).

Le problème d'agrégation est également mentionné par Haining (2009) pour qui l'unité d'agrégation agit tel un filtre dont les valeurs collectées sont dépendantes. L'auteur souligne également qu'avec l'agrégation émerge un problème d'échelle puisque pour différents niveaux d'agrégation spatiale, il existe différents facteurs responsables des résultats observés. Ainsi, même si plusieurs tentatives ont été réalisées afin d'identifier la vraie valeur des coefficients de régression et de corrélation, celle-ci ne peut exister lorsqu'elle est appliquée à des données agrégées : « This outcome is because an "area" such as a census tract is not a random aggregation on individuals, and, unlike "individuals", an "area" is not a well-defined object of study. » (Haining, 2009: 366). Anselin (2002) abonde dans le même sens et rappelle les dangers de l'inférence basée sur des données agrégées (risque d'erreur écologique - ecological fallacy). De plus, le choix de l'unité d'observation est central dans l'analyse des processus spatiaux et peut encourager la présence d'autocorrélation spatiale (Anselin and Griffith, 1988). Finalement, « at an aggregate level, heteroskedasticity should be expected and an i.i.d. [*independent and identically distributed*] assumption for the errors is incompatible with aggregation rule » (Anselin, 2002: 261). Une troisième avenue est de considérer ces observations non répétées en coupes transversales empilées dans le temps (figure 1.8). Cette dernière proposition libère le chercheur des problèmes d'agrégation et des aires modifiables, mais complique la compréhension des processus de dépendance spatio-temporels entre les observations.

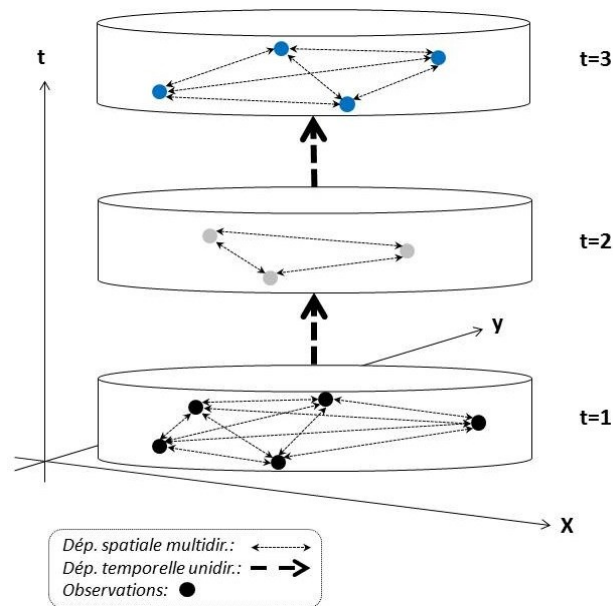


Figure inspirée de Dubé et Legros, 2013

Figure 1.8: Observations en coupes transversales empilées (CTE)

Les propositions méthodologiques afin de conjuguer les dimensions spatiales et temporelles pour les données en coupes transversales empilées sont limitées et peu appliquées dans les travaux empiriques. Il est alors commun de soustraire la dimension temporelle du processus qui structure les observations. En soustrayant celle-ci aux données chronologiquement organisées, le modélisateur suppose une observation simultanée des données et, par conséquent, élimine les contraintes de temps liées aux possibles relations entre les observations (Dubé and Legros, 2013a, Dubé et al., 2012). À ce jour, les données structurées en coupes transversales empilées dans le temps sont généralement traitées avec les méthodes classiques de modélisation spatiale, ce qui suppose un effet nul du temps sur les phénomènes spatialisés. Si ce postulat est supposé vrai (Dubé et al., 2013a, Dubé and Legros, 2013b), il n'est que très rarement vérifié. Cette hypothèse implique donc de nier la

nature dynamique des processus causés par une hétérogénéité spatiale ou structurelle, tel que suggéré par Anselin (2002), et risque de nuire à la validité de certaines statistiques. En effet, selon cette hypothèse, le paramètre autorégressif est biaisé et les conclusions de l'étude sont invalides.

1.4.2 Question et objectifs de recherche

Les analyses empiriques peuvent être de nature confirmatoire ou exploratoire. Ces deux approches proposent souvent des résultats conflictuels en raison de problèmes économétriques notamment causés par la modélisation inadéquate des effets spatiaux (Rey and Janikas, 2005).

La modélisation mathématique tente de schématiser un phénomène empirique afin d'en expliquer les fondements. Le chercheur spécifie alors un *processus générateur de données*⁵ (PGD) à partir duquel il postule les relations entre les variables sélectionnées sur la base du savoir théorique : « The theory will specify a deterministic relationship between the dependent variable and the independent variables » (Greene, 1997: 10). Néanmoins, dans le cadre des sciences sociales, théorie et empirie ne coïncident pas toujours. La définition exacte du PGD demeure donc imprécise : « researchers from social sciences are confronted with substantial specification uncertainty » (Blommestein, 1983: 253). Une partie de cette incertitude résulte du caractère spatialisé des données aux prises avec les problèmes de variables omises et de forme fonctionnelle erronée. Ainsi, dans un contexte spatial ou spatiotemporel, le PGD erroné est susceptible de générer un problème d'autocorrélation.

⁵ De l'anglais *Data Generating Process* (DGP)

La contribution des méthodes économétriques à la prise en compte de la dépendance spatiale est désormais bien établie : « the use of spatial econometric models to address spatial autocorrelation is becoming increasingly standard » (Bhattacharjee et al., 2012: 140). Pourtant, l'inclusion de la dimension temporelle aux modèles de régressions spatiaux demeure un défi pour les données en coupes transversales empilées. C'est pourquoi, afin de contribuer aux avancées des méthodes en économétrie spatiale, la recherche propose d'étudier l'impact de l'omission de la dimension temporelle sur la détection et la correction de l'autocorrélation spatiale pour l'analyse de microdonnées collectées dans le temps et l'espace.

En ce sens, la principale question de recherche soulevée par cette thèse est la suivante: *Quel est l'impact de l'omission de la dépendance temporelle sur la modélisation de données spatiales en coupes transversales empilées ?*

L'intégration de la dimension temporelle par le biais du processus générateur de données (PGD) est définie par le chercheur. Celui-ci peut affecter la mesure de la dépendance spatiale ainsi que sa correction par la spécification d'un modèle autorégressif. La principale question de recherche est décomposée en deux questions spécifiques :

Question spécifique #1 : Quel est l'impact du *processus générateur de données* (PGD) sur la détection de l'autocorrélation spatiale?

Cette question spécifique soulève quatre hypothèses de recherche :

H1a : L'omission du temps surestime les relations spatiales et a un impact sur la détection de l'autocorrélation spatiale (ACS).

H1b : L'utilisation d'une matrice de pondérations spatiales contribue à surestimer les relations spatiales et influence à la hausse la détection de l'ACS lorsque la fonction de distance postulée favorise une forte pondération des relations.

H1c : Contraindre les relations spatiales par le recours à des distances seuils dans la matrice de pondérations spatiales influence la détection de l'ACS. Une distance seuil élevée favorise la présence de relations possiblement abusives surestimant ainsi la mesure de détection de l'autocorrélation spatiale.

H1d : La spécification du modèle économétrique influence la détection de l'autocorrélation spatiale. La forme fonctionnelle ou encore le nombre de variables explicatives pourraient affecter la mesure de la dépendance spatiale.

Les mesures de détection de la présence d'autocorrélation spatiale ne sont habituellement qu'une première étape dont le but ultime est de corriger le problème de manière explicite lors de la modélisation. En ce sens, la seconde question spécifique de recherche est liée à la précédente et en constitue la suite logique.

Question spécifique #2 : Quel est l'impact du *processus générateur de données* (PGD) sur la correction de l'autocorrélation spatiale (modèle autorégressif)?

H2a : L'omission de la dimension temps génère des relations spatiales fallacieuses qui ont un impact sur la mesure des effets de débordements spatiaux (spillover effects).

H2b : Une matrice de pondérations strictement spatiales contribue à surestimer les relations spatiales et biaise à la hausse la mesure des effets de débordements spatiaux (spillover effects).

H2c : Le recours à une distance seuil dans la création de matrices de pondérations spatiotemporelles n'a pas d'impact sur la mesure des effets de débordements spatiaux. L'utilisation d'un produit d'Hadamard des matrices de pondérations spatiales (S) et temporelles (T), permet l'isolement des relations spatiales à une période donnée. Ainsi, l'argument du plus grand mythe de l'économétrie spatiale (LeSage and Pace, 2014a)

appliqué à une matrice spatiale serait également applicable à une telle matrice de pondérations spatiotemporelles.

H2d : La spécification du modèle économétrique influence la mesure des effets de débordements spatiaux. La forme fonctionnelle, la méthode d'estimation ou encore le nombre de variables explicatives pourraient affecter la mesure de la dépendance spatiale.

La recherche vise à rejoindre plusieurs objectifs. Il s'agit dans un premier temps de réaliser une analyse complémentaire aux travaux antérieurs sur le sujet. Ensuite, la recherche propose de mesurer sur la base de données empiriques et des méthodes de méta-analyse (article #1 et #2) les conséquences de l'omission de la dimension temporelle sous-jacente aux données et de la spécification économétrique spatiale sur les statistiques de détection et de correction de l'autocorrélation spatiale. Par la suite, des applications concrètes et actuelles sont réalisées, basées sur les conclusions de l'étude (article #3 et #4 de la thèse). Par cette démarche, le travail tente d'améliorer la compréhension des processus spatiotemporels structurant les données et de contribuer ainsi à l'avancement des connaissances sur l'utilisation de microdonnées dans les méthodes économétriques spatiales.

Plusieurs retombées potentielles sont attendues. Les réflexions méthodologiques permettront de contribuer à l'amélioration des travaux empiriques d'analyse de politiques publiques. Ensuite, la recherche se concentre sur les données immobilières de manières instrumentales et contribue plus largement à la compréhension des phénomènes régionaux sous différentes formes. Tel que mentionné plus tôt, certaines problématiques territoriales impliquent une échelle d'analyse fine et supposent l'utilisation de microdonnées. Les réflexions qui accompagnent ce travail permettront donc d'encourager l'exploitation de telles données au bénéfice de problématiques régionales actuelles tout en outrepassant certaines limites liées aux problèmes d'agrégation, aux problèmes d'erreur écologique et aux aires modifiables (MAUP).

CHAPITRE 2

CADRE THÉORIQUE ET EMPIRIQUE

La thèse porte sur un sujet qui est plus à saveur méthodologique que théorique. Le but est d'explorer l'impact lié à l'omission de la dimension temporelle dans les données spatiales empilées dans le temps pour, ultimement, proposer une approche permettant d'isoler adéquatement la dimension spatiale dans un contexte spatiotemporel dans des applications particulières liées à l'analyse régionale. Plus spécifiquement, les applications et les chapitres porteront sur l'utilisation des microdonnées spatiales empilées dans le temps dont le marché immobilier, associé à l'analyse urbaine, représente un excellent exemple.

Avant de discuter formellement du cadre méthodologique retenu, il semble approprié de présenter l'essentiel du cadre théorique sur lequel les applications immobilières reposent. Les prochaines lignes présentent donc le cadre théorique et abordent les applications empiriques reposant sur le choix d'un modèle statistique et la spécification d'une forme fonctionnelle particulière (ou sur l'idée que se fait le chercheur du processus générateur des données – PGD). C'est sur ce second point que le chapitre introduira plus formellement les particularités méthodologiques de la modélisation économétrique spatiale et spatiotemporelle (Chapitre 3).

Le présent chapitre propose d'aborder la logique de modélisation d'un phénomène ou d'une problématique donnée. Celui-ci est divisé en trois sections. La première est destinée à présenter formellement la théorie hédonique, une perspective largement utilisée dans

l'analyse de données foncières. Par la suite, la seconde section traite de la notion de modélisation, ou de formalisation du concept théorique pour des cas pratiques. Elle est abordée avec une attention particulière aux dimensions propres aux modèles statistiques, au choix de processus générateur des données (PGD) ainsi qu'aux sources des problèmes pouvant survenir dans le cas où les résidus (ou erreurs) du modèle font état de dépendance spatiale. Une courte discussion sur la prise en charge de cette dépendance, par le biais de l'utilisation d'une matrice de pondérations spatiales, est également menée. Finalement, une dernière section présente un schéma conceptuel précisant les intentions de la recherche.

2.1 CADRE THÉORIQUE

2.1.1 Théorie hédonique

Le développement régional ou urbain soulève indubitablement l'application de politiques publiques, dont la mesure des effets peut nécessiter une échelle d'analyse fine. Un moyen d'évaluer des politiques de développement urbain est d'en mesurer des impacts économiques subséquents, notamment via le marché immobilier. Le processus d'évaluation est une étape essentielle du suivi et de l'efficacité de ces politiques mais permettant aussi d'orienter l'application réfléchie des politiques futures.

À titre d'exemple, les aménagements paysagers ou le développement d'un réseau de transport en commun engendrent-ils des retombées autres que sociales ou environnementales? Les valeurs foncières sont reconnues pour fluctuer en fonction de leur localisation, mais également en fonction de ce qui se trouve à proximité. Pour cette raison, le marché immobilier est souvent utilisé pour évaluer l'impact d'un changement dans l'aménagement et, par conséquent, des décisions visant à structurer le développement.

En ce sens, la théorie hédonique, qui repose sur les préférences révélées du marché (acheteurs et vendeurs) représente une façon intéressante d'évaluer l'impact économique de certaines stratégies de développement afin d'orienter les décisions politiques futures.

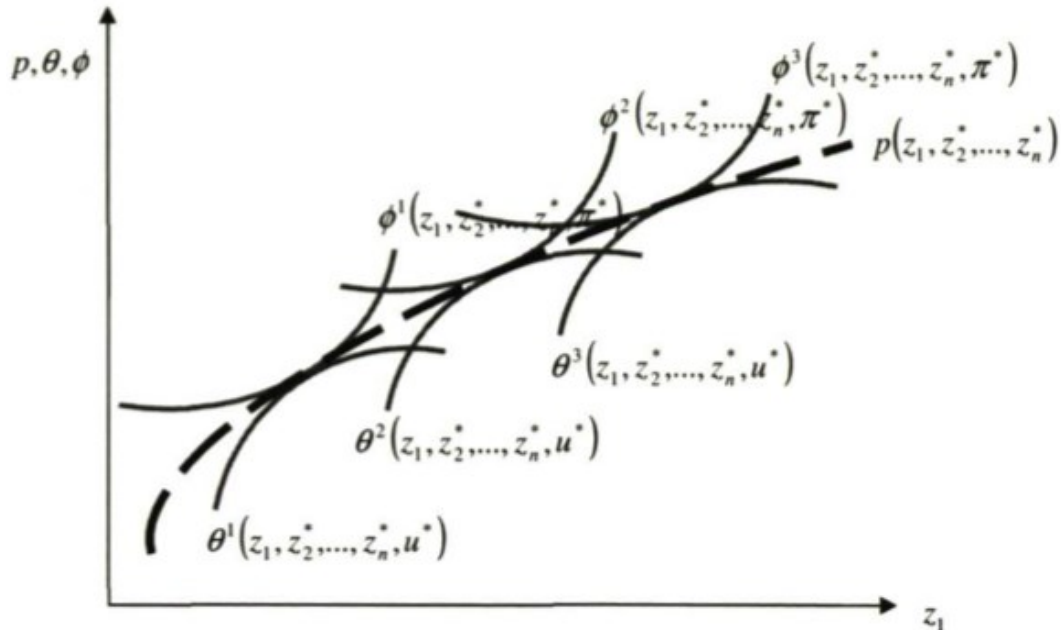
La fonction de prix hédonique est une fonction enveloppe qui illustre l'intersection des fonctions d'enchère (demande) et d'acceptation (offre). La fonction d'enchère ou de valeur, θ , (équation 2.1) illustre le prix maximum que le consommateur est disposé à payer pour un ensemble de combinaisons de caractéristiques et pour un niveau d'utilité donné ainsi que selon son revenu, y , et ses préférences, α (Dubé, 2010) :

$$\theta = \theta(z_1, \dots, z_n, y, \alpha) \quad (2.1)$$

La fonction d'acceptation, Φ , (équation 2.2) est une fonction symétrique à la fonction d'enchère et illustre le prix minimum exigé par un vendeur afin d'accepter de transiger sur le marché un bien, z , composé de n caractéristiques (Maleyre, 1997). Cette fonction intègre un niveau de production, M , et des coûts de production, β (Dubé, 2010):

$$\Phi = \Phi(z_1, \dots, z_n, M, \beta) \quad (2.2)$$

Sur un graphique à deux dimensions, pour lequel seule la caractéristique z_1 varie, alors que les autres paramètres sont fixes ($z_2^*, \dots, z_n^*, u^*, \pi^*$), où u^* correspond au niveau d'utilité lors de la consommation et π^* au niveau de profit du producteur, la fonction de prix hédonique, $p(z_1, z_2^*, \dots, z_n^*)$, permet d'exprimer le prix implicite de marché pour l'attribut z_1 (*ibid.* : 16). Le prix que le consommateur est prêt à payer est donné par la fonction d'enchère $\theta(z_1, z_2^*, \dots, z_n^*, u^*)$, tandis que le prix de vente accepté par le producteur est donné par la fonction d'acceptation $\Phi(z_1, z_2^*, \dots, z_n^*, \pi^*)$. La figure 2.1 illustre cette relation.



Source : Dubé, 2010

Figure 2.1: Détermination du prix implicite pour l'attribut z_1

Une situation d'équilibre intervient lorsqu'une transaction a lieu. À ce moment précis, acheteur et vendeur acceptent de s'entendre sur le prix final du bien. La transaction reflète donc implicitement le processus d'équilibre où l'offre égale la demande, révélant ainsi les préférences du marché. Finalement, Rosen (1974) souligne le fait qu'en étant issus d'une fonction enveloppe, l'équation de prix hédoniques estimés ne représente ni la demande ni l'offre. L'équation de prix ne permet pas, par conséquent, d'identifier la structure des préférences des consommateurs ou de production des producteurs.

L'approche hédonique est définie dans les travaux de Lancaster (1966) et Rosen (1974) : « Hedonic prices are defined as the implicit prices of attributes and are revealed to economic agents from observed prices of differentiated products and the specific amounts

of characteristics associated with them » (Rosen, 1974: 34). Les prix exprimés sont dits implicites car ils ne peuvent être observés qu'indirectement à travers le prix total du bien immobilier et non directement sur le marché (Osland, 2013).

Cette définition renvoie aux précisions apportées par Lancaster sur la théorie du comportement des consommateurs (Debreu, 1959) : « The chief technical novelty lies in breaking away from the traditional approach that goods are the direct objects of utility and, instead, supposing that it is the properties or characteristics of the goods from which utility is derived » (Lancaster, 1966: 133). En résumé, l'utilité qui définit les préférences du consommateur n'est pas reflétée directement par le bien mais plutôt par les caractéristiques du bien (Maleyre, 1997).

En pratique, il n'est en revanche pas possible de connaître avec exactitude la forme de la fonction de prix hédonique, $p(\cdot)$. Celle-ci est habituellement approximée en intégrant un paramètre « ε » qui tient compte des erreurs de mesures associées à l'omission de caractéristiques présentes dans la fonction $p(\cdot)$. La fonction de prix hédonique est souvent estimée en recourant à la régression linéaire multiple : les coefficients obtenus permettent ainsi de reconstruire la valeur implicite moyenne de chacun des attributs, autant physiques qu'environnementaux (équation 2.3).

$$P_i = \hat{p}(z_{1i}, \dots, z_{ni}, \varepsilon_i) \quad (2.3)$$

Ces prix implicites sont obtenus par la régression de la valeur immobilière sur un ensemble d'attributs du logement (Cavailhès, 2005). Ainsi les prix hédoniques réfèrent à la valeur implicite des attributs considérés (Can, 1990) qui ne font pas directement ou individuellement l'objet d'échanges sur le marché (Maleyre, 1997). Ces attributs peuvent être de nature intrinsèque à la propriété vendue ou environnementale, c'est-à-dire liés à sa localisation (Dubé et al., 2011a).

L'approche des prix hédoniques se heurte à certaines limites. Rosen (1974) souligne le fait qu'en étant issus d'une fonction enveloppe, les prix hédoniques estimés ne représentent ni la demande ni l'offre mais davantage une combinaison des deux. Ils ne peuvent par conséquent pas à eux seuls identifier la structure des préférences des consommateurs ou de production des producteurs. Il en découle une difficulté à discerner des facteurs d'offre et de demande dans la fonction hédonique rendant difficile l'interprétation des prix qui en sont dérivés (Des Rosiers, 1991).

Évidemment, les défis sur l'application de l'approche hédonique sont nombreux : choix de la forme fonctionnelle, possibilité d'effets non linéaires, mesure et contrôle de l'ensemble des variables significatives et contrôle des hypothèses de base sur le comportement des résidus. L'ensemble de ces facteurs peut se résumer aux défis liés au processus générateur des données postulé par le chercheur afin de respecter les hypothèses de base du modèle de régression linéaire multiple.

2.2 CADRE EMPIRIQUE

2.2.1 Le modèle de régression linéaire multiple (RLM) standard, MCO

L'approche hédonique repose sur l'estimation d'un modèle de régression linéaire multiple (RLM) destiné à expliquer la variation d'une variable dépendante, y , telle qu'un prix de vente, en fonction d'un ensemble de variables explicatives (x_1, x_2, \dots, x_k) , telles que des caractéristiques d'une propriété immobilière. Il convient d'en rappeler les fondements avant de décrire les défis associés à la spécification du modèle.

$$y_i = f(x_{i1}, x_{i2}, \dots, x_{ik}) + \varepsilon_i \quad (2.3)$$

La RLM constitue l'outil le plus utile à l'économètre (Greene, 1997). Sous sa forme générique, le modèle de régression linéaire multiple exprime une relation, linéaire dans les paramètres $(\beta_1, \beta_2, \dots, \beta_k)$, entre les variables sélectionnées :

$$y_i = x_{i1}\beta_{i1} + x_{i2}\beta_{i2} + \dots + x_{ik}\beta_{ik} + \varepsilon_i \quad (2.4)$$

où le terme β_k correspond au coefficient témoignant de la force et du sens de la relation entre une variable dépendante et une variable indépendante associée, x_{ik} pour une observation i donnée.

Puisqu'il est en pratique impossible d'observer une relation linéaire parfaite, le terme d'erreur, ε_i , permet de capter et d'intégrer les effets non mesurés et non pris en compte par les variables indépendantes. Le terme d'erreur synthétise trois types d'erreurs : i) l'erreur de spécification qui correspond à l'omission de variables explicatives; ii) l'erreur de mesure qui corrige le fait que les données ne représentent pas parfaitement le phénomène observé et iii) l'erreur d'échantillonnage puisque les observations diffèrent d'un échantillon à l'autre impliquant que les estimations varient également légèrement (Bourbonnais, 2003).

Afin d'estimer le coefficient $\hat{\beta}_k$, associé à une variable dépendante x_{ik} , la méthode des moindres carrés ordinaires (MCO) est habituellement appliquée dans le cadre d'une RLM⁶. Le principe des MCO est d'obtenir une estimation pour les paramètres d'une droite de régression sur un plan cartésien de sorte que soit minimale la distance, au carré, entre la droite et l'ensemble des points (observations) individuels. En d'autres termes, il s'agit de minimiser les erreurs au carré ($Min \sum_{i=1}^n \varepsilon_i^2$). D'autres méthodes d'estimation existent, mais la procédure MCO permet l'obtention d'estimateurs *BLUE* (*Best Linear Unbiased*

⁶ Le coefficient β est noté $\hat{\beta}$ après l'estimation. De manière générale on peut également noter le coefficient β_{MCO} dans le cadre d'une estimation par la méthode des MCO.

Efficiency Estimators) lorsque les hypothèses sont respectées. Les estimateurs sont donc sans biais et de variance minimale. La présence de dépendance spatiale et/ou temporelle est susceptible d'invalider l'hypothèse d'indépendance du terme d'erreur (Hypothèse 4) et d'empêcher l'estimation d'un modèle *BLUE*.

Le modèle de RLM doit respecter certaines hypothèses. Deux types d'hypothèses existent : les hypothèses stochastiques et les hypothèses structurelles. Les premières portent sur le terme d'erreur, tandis que les secondes réfèrent à l'organisation et aux relations entre les variables considérées (tableau 2.1).

Tableau 2.1: Les hypothèses du modèle de RLM (*Bourbonnais, 2003*)

HYPOTHESE STOCHASTIQUES	
Hypothèse 1	Les valeurs x_i sont observées sans erreur et sont exogènes ;
Hypothèse 2	$E(\varepsilon_i) = 0$, l'espérance mathématique de l'erreur est nulle. Autrement dit, les erreurs sont de moyenne nulle. Dans le cas contraire, la constante du modèle est biaisée mais pas les paramètres estimés ;
Hypothèse 3	$E(\varepsilon_i^2) = \sigma_\varepsilon^2$, homoscédasticité : la variance de l'erreur est constante ;
Hypothèse 4	$E(\varepsilon_i \varepsilon_{i'}) = 0$, les erreurs sont non corrélées (indépendantes) entre les observations. L'erreur associée à une observation n'est pas autocorrélée avec l'erreur d'autres observations ;
Hypothèse 5	$Cov(x_i, \varepsilon_i) = 0$, l'erreur est indépendante des variables explicatives. Il n'y a donc pas de corrélation entre les régresseurs et le terme d'erreur.
HYPOTHESES STRUCTURELLES	
Hypothèse 6	Absence de colinéarité entre les variables explicatives. Il n'y a donc pas d'autocorrélation entre les régresseurs ;
Hypothèse 7	$(X'X)/n$ tend vers une matrice finie non singulière (donc inversible) ;
Hypothèse 8	$n > k + 1$, le nombre d'observations est supérieur au nombre de variables explicatives.

En utilisant les hypothèses de la RLM (tableau 2.1), il est possible de démontrer⁷ que les estimateurs $\hat{\beta}$ sont des estimations sans biais de β , ($E(\hat{\beta}|X) = \beta$). Autrement dit, l'espérance de l'estimateur des MCO coïncide avec la *vraie* valeur du paramètre. Il est également possible de démontrer, sous l'hypothèse de normalité des erreurs (hypothèse 3), que la variance du coefficient estimé est la meilleure parmi tous les estimateurs linéaires non biaisés β , ($V(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$). La variance des estimateurs est qualifiée d'efficace. Puisque les estimateurs MCO sont sans biais pour le modèle de RLM, ils le sont également pour les échantillons de taille infinie. On dira qu'ils sont sans biais de manière asymptotique. Il est aussi démontré que la matrice de variance-covariance de $\hat{\beta}$ se rapproche de 0 lorsque la taille de l'échantillon tend vers l'infini (Kennedy, 2003).

Évidemment, une partie de la validité des hypothèses repose sur le fait que le modèle linéaire est spécifié de manière complète; autrement dit, que l'équation de prix reflète le processus générateur des données (PGD) pour la variable dépendante.

2.2.2 Les défis du processus générateur de données (PGD)

Le modèle de RLM repose indubitablement sur la spécification d'un processus générateur de données (PGD). La figure 2.2 illustre la démarche de modélisation.

⁷ Voir Bourbonnais (2003) pour le détail de la démonstration.

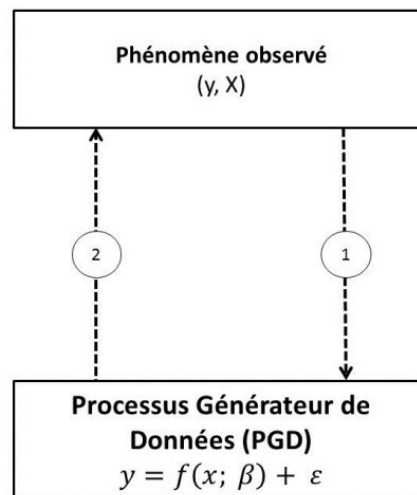


Figure 2.2: Modélisation en contexte aspatial

Le PGD est d'abord formulé sur une base théorique : « The theory will specify a deterministic relationship between the dependent variable and the independent variables » (Greene, 1997: 10). Néanmoins, dans plusieurs domaines des sciences sociales, les théories ne résultent que rarement en des fonctions statistiques permettant une estimation directe. Il n'est pas courant d'avoir directement la forme exacte de la relation unissant la variable dépendante et les variables indépendantes. La définition exacte du PGD demeure donc inconnue : « researchers from social sciences are confronted with substantial specification uncertainty » (Blommestein, 1983: 253).

Au-delà de la relation fonctionnelle entre les variables indépendantes et expliquées, la validité de l'interprétation du modèle économétrique dépend du respect des hypothèses de base liées au comportement des termes d'erreurs de la RLM. Sheppard (1999) déplore la tendance des économistes à se concentrer sur l'optimisation de la méthode d'analyse et à négliger la structure et la qualité des données incluses dans la modélisation.

L'autocorrelation du terme d'erreur rend inefficace la variance des coefficients estimés, créant un accroissement du risque d'erreur de type 1⁸ et de la statistique d'ajustement du modèle (R^2) (Anselin, 2002, Haining, 2009). Les statistiques classiques comme les tests de significativité des coefficients deviennent alors biaisées : « This is problematic since failing to account for serial and spatial autocorrelation when present causes the OLS estimators to lose its property of efficiency » (Elhorst, 2001: 2001).

Il existe deux sources principales d'autocorrélation des résidus du modèle de régression. La première renvoie à une mauvaise spécification économétrique, tandis que la seconde se rapporte aux résultats de la modélisation explicite de l'espace (Anselin and Griffith, 1988, Le Gallo, 2002).

Une partie de l'incertitude associée à la spécification du PGD résulte (i) du choix de la forme fonctionnelle et (ii) de la nature des attributs du logement à inclure dans MPH (Osland, 2013). Lors d'une modélisation hédonique, la forme de la relation entre une variable dépendante et une variable indépendante est inconnue. Le chercheur se base donc sur la théorie et une certaine intuition pour postuler la forme de relation entre les variables qui peut être de nature linéaire, semi-logarithmique ou double-logarithmique.

La forme linéaire du MPH ne nécessite aucune transformation. Une variation dans la valeur d'une variable indépendante engendrera un changement constant dans la variable dépendante. Néanmoins, les attributs qui composent les biens complexes tels que les logements ne respectent pas nécessairement cette intuition en raison notamment d'un coût fixe de production de sorte que le prix unitaire n'est pas constant (Cavailhès, 2005). En clair, le fait de choisir une troisième chambre lors de l'achat d'une propriété ne fait pas croître son prix total d'une valeur identique à l'ajout d'une seconde. Ensuite, la forme semi-

⁸ L'erreur de type 1 se caractérise par un rejet abusif de l'hypothèse nulle lors d'un test statistique. On postulera par exemple injustement l'existence d'une relation significative entre deux variables.

logarithmique permet de normaliser la distribution de la variable dépendante. Une transformation logarithmique est alors appliquée sur le prix de vente final de la résidence. Enfin, la forme double-logarithmique implique que le modèle comporte à la fois une transformation logarithmique de la variable dépendante mais également dans sa partie déterministe (variables explicatives). De manière générale, ces problèmes de spécification peuvent être identifiés par des tests statistiques et résolus par une transformation mathématique simple.

2.2.3 Intégration de l'espace et du temps au PGD

L'utilisation de données géoréférencées soulève également le défi de l'inclusion de la dimension spatiale. L'autocorrélation spatiale peut émerger (i) d'une mauvaise spécification du modèle causée par l'omission de variables (McMillen, 2003, McMillen, 2010); (ii) par l'utilisation d'une mauvaise forme fonctionnelle ou (iii) par des erreurs de mesure (Le Gallo, 2002).

Le problème de l'omission de variables se présente de manière croissante depuis les 20 dernières années (Osland, 2013). Ces variables omises ne se rapportent généralement pas directement au bien transigé mais à la définition de son environnement et de son voisinage (Cavailhès, 2005, McMillen, 2010) : « The main concern has been about omitted or unobserved local or regional characteristics rather than attributes related to the house itself » (Osland, 2013: 65). Anselin et Arribas-Bel (2011) constatent également la récurrence de cette lacune qu'ils associent à l'omission de variables contextuelles : « In empirical applications, a common problem is the presence of unobserved local or regional variables that may give rise to spatial error correlation » (*ibid.* : 1). Ces variables omises peuvent être spatialement corrélées et engendrer l'estimation de résidus hétéroscédastiques, invalidant les résultats du modèle économétrique (Anselin and Arribas-Bel, 2011).

Ensuite, la formalisation dans le PGD des relations spatiales hautement non linéaires constitue un défi en soi, selon McMillen (2010). L'auteur rappelle la volonté des études empiriques de mesurer l'effet de la proximité sur les valeurs marchandes du sol à partir de la perspective du modèle Alonso-Muth-Mills, selon lequel à mesure que l'on s'éloigne du centre-ville cet effet diminue. L'effet varierait donc en fonction de la distance. En pratique, la forme de la fonction liant distance et prix demeure inconnue. Au final, « the main difficulty with modeling spatial data is the combination of non-linearity and missing variables that are correlated over space » (*ibid.*: 138).

D'autres effets de dépendance spatiale de nature plus structurelle sont à considérer. En effet, les données peuvent être soumises à un processus commun, attribuable à des facteurs contextuels, qui met les observations en relation et qui les structure selon une organisation particulière (phénomène de diffusion, par exemple). Il s'agit alors pour le modélisateur d'appréhender ces processus qui intègrent deux formes de dépendance spatiale distinctes. La première est caractérisée par des effets de voisinage correspondant à une internalisation commune de facteurs environnementaux susceptibles d'influencer les valeurs des observations en un lieu donné (Ham et al., 2012). Cette instabilité structurelle spatiale est nommée hétérogénéité spatiale et occasionne des problèmes économétriques tels que la présence d'hétéroscédasticité ou une variation aléatoire de coefficients. Il s'agit d'une situation commune dans l'analyse de phénomènes régionaux (Anselin and Griffith, 1988). La seconde se présente à travers des effets de débordements, c'est-à-dire par l'internalisation pour une observation de caractéristiques d'observations voisines (Can, 1992, Kestens et al., 2006). Ainsi, dans le cadre de valeurs résidentielles, le prix d'une maison est fonction du prix des maisons voisines (Can, 1990).⁹

⁹ Des Rosiers et al. (2011) misent sur les effets de pairs (*peer effects*) pour contrôler les effets d'une autocorrélation spatiale endogène associée à des effets de débordements. L'inclusion des prix moyens d'un secteur donné comme variable explicative contribuerait ainsi à reproduire les effets d'un modèle de type SAR.

En résumé et de manière formelle, la dépendance spatiale peut donc être associée à une mauvaise spécification du modèle ou à la modélisation explicite de l'espace tel que présenté aux équations 2.5 et 2.6 :

$$y_i = f(y_j, x_{1i}, \dots, x_{Ki}, x_{1j}, \dots, x_{Kj}, \beta_k, \rho, \theta_k, \varepsilon_i) \quad (2.5)$$

ou encore:

$$\varepsilon_i = g(\varepsilon_j, \lambda, \xi_i) \quad (2.6)$$

où y correspond à la variable dépendante respectivement associée à une observation i et j ; x_k représente les k variables explicatives; ρ , θ et λ sont des vecteurs de paramètres associés aux coefficients autorégressifs des variables respectivement dépendante, indépendante et du terme d'erreur. Ces paramètres autorégressifs permettent d'établir l'effet moyen des relations de dépendance spatiales sur une observation donnée. Finalement, ε et ξ sont des termes d'erreurs. Le premier est autocorrélé et ne respecte donc pas l'hypothèse d'indépendance (tableau 2.1). En revanche, le second respecte cette hypothèse. La forme fonctionnelle de f et g exprime de quelle manière la réalisation de y et ε en un point dans l'espace est associée à leurs valeurs en d'autres lieux.

La relation entre les variables dépendantes ou les termes d'erreurs pour chaque paire d'observations est formalisée par une matrice de pondérations qui représente les interactions entre différentes positions géographiques. Cette matrice de pondérations spatiales est spécifiée de manière exogène par le chercheur et affecte directement le comportement du

Néanmoins cette approche s'expose aux problèmes associés à la délimitation optimale du secteur (*Modifiable Areal Unit Problem, MAUP* – Openshaw, 1988).

terme d'erreur. C'est pourquoi, dans le cadre de la présente recherche elle est inscrite dans la logique du processus générateur de données.

2.3 DE LA THÉORIE À LA PRATIQUE : SCHÉMA CONCEPTUEL

Les valeurs résidentielles constituent un exemple probant de microdonnées spatiales empilées dans le temps et dont les résultats sont inscrits dans un contexte territorial où les distances sont à la fois spatiales et temporelles. Il y a désormais un consensus sur l'importance de la localisation dans la détermination des valeurs foncières (distance spatiale). L'influence de l'espace se présente dans la détermination de ces valeurs à travers des caractéristiques de voisinage, d'accessibilité ou environnementales (Bramley et al., 2008, Des Rosiers, 2002, Hite et al., 2001, Mohammad et al., 2013). La proximité spatiale avec diverses aménités ou d'autres biens semblables peut ainsi affecter la détermination du prix de vente et tendre à entretenir la présence de structures spatiales particulières. Les effets spatiaux sont généralement contrôlés par l'inclusion de variables spatiales diverses. Néanmoins, il est désormais reconnu que les relations spatiales ne sont pas homogènes (ou isotropiques) à travers l'espace (Anselin, 1990, Anselin, 1998, McMillen, 2010) et qu'une spécification erronée des modèles économétriques affecte la validité des estimateurs.

La distance temporelle est également susceptible d'influencer la variation des valeurs. En effet, certains facteurs dynamiques peuvent influencer les prix de vente. Pensons notamment aux contextes économique, politique ou encore social. En revanche, une faible fréquence de répétition des transactions dans le temps rend l'analyse des valeurs immobilières méthodologiquement difficile :

Housing prices are well known to be influenced by the prices of recent house sales nearby, one must allow for possible spatio-temporal dependencies between such prices [...] In addition there are generally a host of other

processes occurring over time that result in unobserved temporal autocorrelations among housing prices. But since individual housing sales do not occur at regular time intervals, it is difficult to model such processes in terms of standard discrete time series. (Smith and Wu, 2009: 334).

L'un des défis de l'analyse de telles données en coupes transversales empilées dans le temps réside dans la spécification du modèle économétrique adéquat pour assurer le respect des hypothèses de base sur le comportement des résidus du modèle, plus particulièrement de la prise en compte de la dimension spatiale dans un contexte spatiotemporel.

Pour faire suite à la problématique que nous avons développée, le schéma conceptuel (figure 2.3) met en relation le processus générateur de données (PGD), soit la formalisation de la relation empirique observée et des relations spatiotemporelles, avec la détection/correction de la dépendance spatiale.

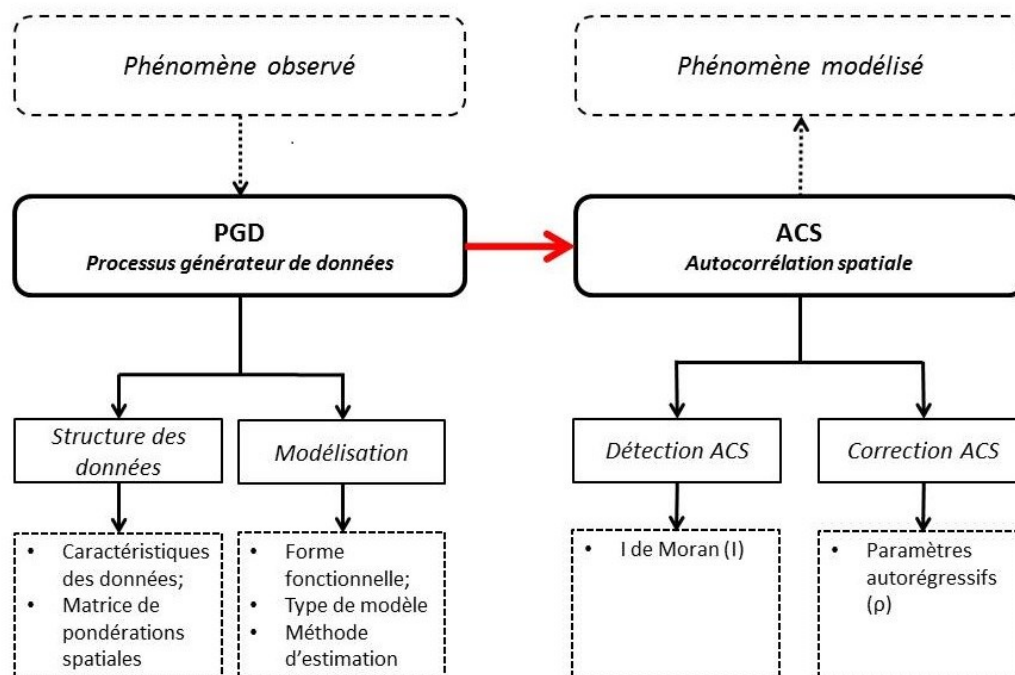


Figure 2.3: Schéma conceptuel

Dans le cadre d'une analyse régionale, un phénomène observé est inscrit dans l'espace et caractérisé par la présence d'une forme de dépendance spatiale structurant les observations et leurs interactions. Le PGD permet de formaliser cette structure. Plus précisément, il permet d'établir les bases de la démarche de modélisation de gérer la présence d'autocorrélation spatiale qui a été présentée comme la mesure de la dépendance spatiale. Le PGD permet donc d'en détecter les traces et d'en corriger les effets afin d'interpréter plus justement un phénomène inscrit dans l'espace.

Le prochain chapitre est consacré à l'aspect méthodologique de la recherche. Les éléments plus formels propres à la modélisation économétrique spatiale y sont présentés : les démarches de modélisation, les tests de détection de l'autocorrélation spatiale et les modèles destinés à corriger la présence de dépendance spatiale.

CHAPITRE 3

MÉTHODOLOGIE

Les premiers chapitres de la thèse ont mis en évidence la contribution des microdonnées dans l'analyse régionale ainsi que les particularités de leur inclusion dans des dynamiques spatiales. Le principe d'autocorrélation spatiale a été présenté et la logique des méthodes économétriques spatiales destinées à en contrôler les effets et à garantir la validité des résultats de recherche a été introduite.

Par la suite, le lecteur a été initié aux bases de la modélisation hédonique et les principes de la régression linéaire multiple ont été rappelés. Une attention particulière a été portée sur les défis associés à la formulation d'un PGD destiné à formaliser mathématiquement un phénomène empirique. À ce stade, la dimension spatiale est susceptible d'introduire plusieurs biais liés à la forme fonctionnelle des relations, à l'omission de variables ou à leur structure dans l'espace.

Le présent chapitre est de nature méthodologique. Il est composé de trois sections portant respectivement sur la modélisation des phénomènes spatiaux, la modélisation des phénomènes spatio-temporels et les principes de la méta-analyse (MA). La structure spatiale des observations jouant un rôle central, la première section du chapitre se concentre sur la formalisation de cette structure par la spécification d'une matrice de pondérations spatiales. La démarche de modélisation ainsi que la statistique du I de Moran destinée à détecter la présence d'autocorrélation spatiale sont d'abord abordées. En cas de traces de dépendance spatiale, les stratégies de correction de ces effets sont décrites. La première repose sur l'estimation d'un modèle incluant des effets fixes, tandis que la deuxième propose de

recourir à une spécification autorégressive (modèle autorégressif spatial, SAR, et modèle d'erreur spatial, SEM).

À la suite de la présentation des rouages de la modélisation économétrique spatiale, la question de la prise en compte de la dimension temporelle est abordée. Si quelques approches différentes sont considérées, la discussion porte principalement sur la formalisation d'une matrice de pondérations spatiotemporelles générale.

Finalement, la recherche propose de réaliser une méta-analyse (MA) afin d'expliquer les variations observées dans les statistiques spatiales. Dans un premier temps, la notion de méta-analyse est définie. Elle est suivie d'une description des principes d'effet de taille (ET) et de méta-régression (MR). Finalement les limites de la MA sont discutées.

3.1 MODÉLISATION DES PROCESSUS SPATIAUX

Inscrits dans l'espace, les phénomènes régionaux sont largement influencés par les dynamiques spatiales. Rappelons que l'économétrie spatiale permet d'analyser ces phénomènes en se reposant sur la première loi de la géographie attribuée à Tobler (1979) postulant que tous les éléments inscrits dans l'espace sont reliés, mais que les liens qui les unissent varient en fonction de la distance. Les éléments les plus proches entretiennent plus de liens, et du même fait, des relations de dépendance plus fortes.

Avant de poursuivre la discussion vers certains principes généraux de la modélisation économétrique spatiale, il convient de s'attarder aux matrices de pondérations spatiales dont la vocation est de formaliser ces relations si importantes.

3.1.1 La matrice de pondérations spatiales (S)

La matrice de pondérations spatiales permet de postuler des relations qui structurent les observations. LeSage et Pace (2014a et 2014b) remettent en question l'impact de cette matrice et vont jusqu'à la présenter comme le plus grand mythe de l'économétrie spatiale. La matrice de pondérations demeure la pierre angulaire des méthodes économétriques spatiales et constitue pour d'autres sa faiblesse : « The specification of the weights matrix is a matter of some arbitrariness and is often cited as a major weakness of the lattice approach » (Anselin, 2002: 257). Bhattacharjee et al. (2012), Elhorst (2010) et Chen (2009 dans Chen, 2012) abondent dans le même sens. En raison des interactions spatiales et de la distribution hétérogène des observations, les phénomènes spatiaux entretiennent des relations hautement non linéaires (McMillen, 2010) que la matrice de pondérations spatiales doit permettre de capter au risque de nuire aux analyses : « If we fail to choose a proper weight function, the result and effect of spatial analysis may be no satisfying and convincing, or even the calculation will be distorted » (Chen, 2012: 66). Ainsi, la réalisation des statistiques de détection de la dépendance spatiale et l'estimation des modèles autorégressifs en est garant.

La matrice de pondérations spatiales repose sur la définition d'une forme de proximité entre les observations. Celle-ci peut être interprétée en termes de contiguïté lorsque les observations sont des polygones (des régions par exemple)¹⁰ ou, en termes de distance lorsque les observations sont des points géolocalisés par des coordonnées (latitude, longitude). La distance entre les points peut se calculer par l'application du théorème de

¹⁰ La matrice binaire peut être appliquée dans le cas de points géolocalisés par les méthodes des polygones de Thiessen ou de Delaunay.

Pythagore¹¹. Sur la base de ces relations, une matrice de pondérations spatiales \mathbf{S} de dimensions $[N \times N]$ est créée où s_{ij} représente la relation entre les observations i et j (équation 3.1) et N représente le nombre total d'individus (ou observations) considéré. Afin de souligner l'importance de la proximité, l'inverse de la distance peut être utilisée et une distance seuil postulée, \bar{d} :

$$s_{ij} = \begin{cases} g(d_{ij}) & \text{si } d_{ij} \leq \bar{d}_i \\ 1 & \text{si } d_{ij} = 0 \quad \forall i \neq j \\ 0 & \text{sinon} \end{cases} \quad (3.1)$$

Sans perte de généralité, la pondération spatiale est fonction de la distance entre les observations ($g(d_{ij})$). Cette fonction de distance peut prendre plusieurs formes, mais trois types de transformations sont particulièrement utilisés. La première spécification repose sur une fonction inverse de la distance séparant deux observations, $1/d_{ij}^\alpha$, où α est un paramètre de pénalité permettant de mettre un poids plus ou moins important sur les distances plus faibles. Cette transformation octroie un poids important aux observations proches avant de décroître fortement avec la distance. La seconde transformation suggère une fonction exponentielle négative de la distance, $\exp(-d_{ij})$, et permet quant à elle d'imputer un poids moins élevé aux observations proches tout en fournissant une décroissance moins rapide de cette valeur avec la distance. Il est également commun d'utiliser une spécification binaire des pondérations afin d'identifier les observations qui ont une relation spatiale de proximité (oui ou non). Dans ce cas, toutes les observations postulées comme voisines posséderont un poids identique et les éléments de la matrice de pondérations spatiales prendront une valeur de 1. Dans tous les cas, la pratique suggère habituellement d'utiliser une limite spatiale pour délimiter les relations, à partir d'un certain rayon (\bar{d}_i), afin d'éviter le problème de sur-connectivité de la matrice (Smith, 2009).

¹¹ La distance entre les observations i et j , notée d_{ij} , est déterminée par :

$$d_{ij} = \sqrt{(lat_i - lat_j)^2 + (long_i - long_j)^2}$$

Tel que présenté à l'équation 3.2, chaque observation est représentée sur chaque ligne et chaque colonne de la matrice de pondérations spatiales. À l'intersection de celles-ci se trouvent l'élément s_{ij} illustrant la relation entre l'observation i et l'observation j non nulle lorsqu'un lien spatial est postulé entre celles-ci.

$$\mathbf{S} = \begin{bmatrix} 0 & s_{12} & \cdots & s_{1j} & \cdots & s_{1n} \\ s_{21} & 0 & \cdots & s_{2j} & \cdots & s_{2n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{i1} & s_{i2} & \cdots & s_{ij} & \cdots & s_{in} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nj} & \cdots & 0 \end{bmatrix} \quad (3.2)$$

Les deux principales propriétés de la matrice de pondérations spatiales sont les suivantes (Chen, 2013) : 1) : la matrice \mathbf{S} est carrée et symétrique de sorte que $s_{ij} = s_{ji}$; 2) Il est admis par convention qu'une observation ne peut être sa propre voisine, ainsi les éléments de la diagonale de la matrice de pondérations sont nuls, $s_{ii} = 0$. Par ailleurs, afin de faciliter la comparaison des pondérations, il est commun de standardiser (*normaliser*) en ligne la matrice \mathbf{S} de sorte que la somme des éléments de chaque ligne soit égale à 1, $s_{ij}^* = s_{ij} / \sum_{j=1}^N s_{ij}$. Il est à noter qu'après la standardisation de la matrice de pondérations, la première propriété ne tient en partie plus; la matrice demeure carrée mais n'est plus symétrique.

La matrice de pondérations spatiales joue un rôle central dans l'interprétation des processus spatiaux. Néanmoins, il faudrait relativiser son impact sur les résultats de recherche selon LeSage et Pace (2014a). En effet, les auteurs suggèrent que si certaines études concluent à un impact important de la spécification de la matrice sur les résultats d'estimations, trop souvent les praticiens omettent de considérer un possible problème de spécification du modèle comme responsable. Ensuite, les résultats d'estimations pourraient faire l'objet d'interprétations erronées de la part des chercheurs. Ainsi, LeSage et Pace

(2014a, 2014b) semblent accuser une compréhension générale des méthodes économétriques par les chercheurs. Ainsi, les sources de variations des résultats ne sont pas uniquement associées à la formalisation des relations spatiales mais à un ensemble de facteurs. À ce titre, l'omission de la dimension temporelle devrait être considérée comme une source supplémentaire potentielle d'impact sur les estimations.

Les prochaines sections du chapitre cibleront d'abord la démarche de modélisation dans un contexte strictement spatial. Par la suite, la problématique de l'addition de la dimension temporelle sera présentée.

3.1.2 Démarches de modélisation et tests de détection des processus spatiaux

L'économétrie spatiale propose un ensemble de modèles statistiques destinés à modéliser les phénomènes en considérant la présence d'autocorrélation. Pour ce faire, une des démarches préconisées consiste à estimer un modèle par moindres carrés ordinaires (MCO) et à tester statistiquement la présence d'autocorrélation spatiale entre les résidus du modèle statistique afin de valider si les hypothèses d'indépendance du modèle de régression linéaire multiple sont respectées (approche du particulier au général). Une seconde approche vise à estimer un modèle autorégressif complet intégrant toutes les formes de dépendances spatiales possibles afin d'en tester statistiquement la significativité (approche du général au particulier). L'objectif est de valider la spécification d'un PGD adéquat à la modélisation du phénomène observé (figure 3.1).

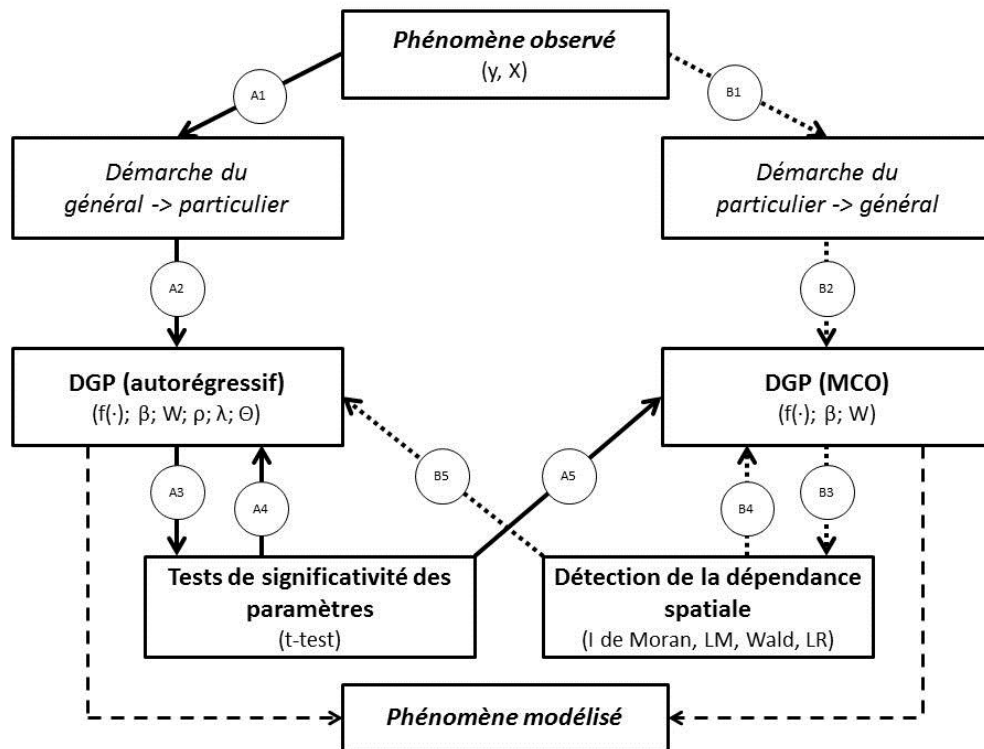


Figure 3.1: Modélisation des processus spatiaux

Dans la première démarche (du particulier au général, relation B de la figure 3.1), le phénomène observé mène à la spécification d'un PGD selon le principe proposé pour une analyse aspatiale qui sera estimée par la méthode des MCO (relation B2). La présence de dépendance dans la structure spatiale des résidus de la régression est ensuite testée (relation B3). Le I de Moran (Moran, 1950, Moran, 1948) est la statistique la plus utilisée pour

évaluer la présence d'autocorrélation spatiale (Griffith, 2012, Le Gallo, 2002).¹² Il permet de quantifier les patrons spatiaux présents dans les données collectées afin de déterminer si celles-ci sont aléatoirement réparties dans l'espace (Haining, 2009). Il exprime l'importance de la différence des valeurs entre toutes les paires d'observations contiguës (Caloz and Collet, 2011) ou proches (équation 3.3).¹³

$$I = \frac{N}{S_0} \left(\frac{\tilde{\varepsilon}' \mathbf{S} \tilde{\varepsilon}}{\tilde{\varepsilon}' \tilde{\varepsilon}} \right) \quad (3.3)$$

où N représente le nombre d'observations; S_0 la somme des éléments de la matrice de pondérations ($S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij}$); $\tilde{\varepsilon}$ la variable de résidus du modèle de régression MCO; $\mathbf{S} \tilde{\varepsilon}$ la variable spatialement décalée basée sur une matrice de pondérations spécifiée de manière exogène, \mathbf{S} . Lorsque cette dernière est standardisée en ligne, la somme de chaque ligne est égale à 1 et $S_0 = N$. L'équation 3.3 peut alors être simplifiée :

$$I = \frac{\tilde{\varepsilon}' \mathbf{S} \tilde{\varepsilon}}{\tilde{\varepsilon}' \tilde{\varepsilon}} \quad (3.4)$$

¹² Le Gallo (2002) fait également remarquer que si l'autocorrélation spatiale peut résulter d'une spécification économétrique erronée, alors l'indice I de Moran peut être utilisé comme un indicateur général de spécification.

¹³ Contrairement à l'indice de Moran qui permet une mesure de l'autocorrélation spatiale globale, la statistique LISA (*Local Indicators of Spatial Autocorrelation*) en permet une mesure locale. Ainsi, même si

La statistique I de Moran exprime le rapport entre la covariance pondérée pour des unités voisines et la variance totale observée (Le Gallo, 2002: 143). La standardisation en ligne permet de simplifier son interprétation puisque l'indice d'autocorrélation calculé est borné à un domaine de valeurs entre -1 et 1, où 0 représente l'absence d'autocorrélation spatiale. Le test statistique pour la présence d'autocorrélation spatiale (H_0 : *absence d'autocorrélation spatiale*) est basé sur une statistique centrée réduite et normalement distribuée, $N \sim (0,1)$ (équation 3.5).

$$z = \frac{I - E(I)}{\sqrt{\sigma_I^2}} \quad (3.5)$$

où I représente l'indice de Moran; $E(I)$ son espérance et σ_I^2 sa variance¹⁴.

D'autres statistiques permettent d'identifier le type de processus autorégressif présent dans les données (multiplicateur de Lagrange, test de Wald et ratio de vraisemblance) qui peut porter sur la variable dépendante (*Spatial Autoregressive Model*, SAR) ou sur le terme d'erreur (*Spatial error model*, SEM). Le multiplicateur de Lagrange (*LM*) est un test contraint permettant d'orienter le chercheur vers l'un ou l'autre de ces processus. Par son caractère contraint, le test d'hypothèse permet d'évaluer si les paramètres autorégressifs (respectivement ρ ou λ) sont différents ou non de 0.

Le test de Wald (W) est un test non contraint et repose sur l'estimation des modèles autorégressifs. Le test d'hypothèse n'est donc pas associé à la valeur nulle du paramètre

aucune autocorrélation spatiale globale significative n'est détectée, des traces de dépendance spatiale locales peuvent exister.

¹⁴ Voir Dubé et Legros (2014) pour plus de détails sur le calcul de la variance et de l'espérance pour la statistique de Moran.

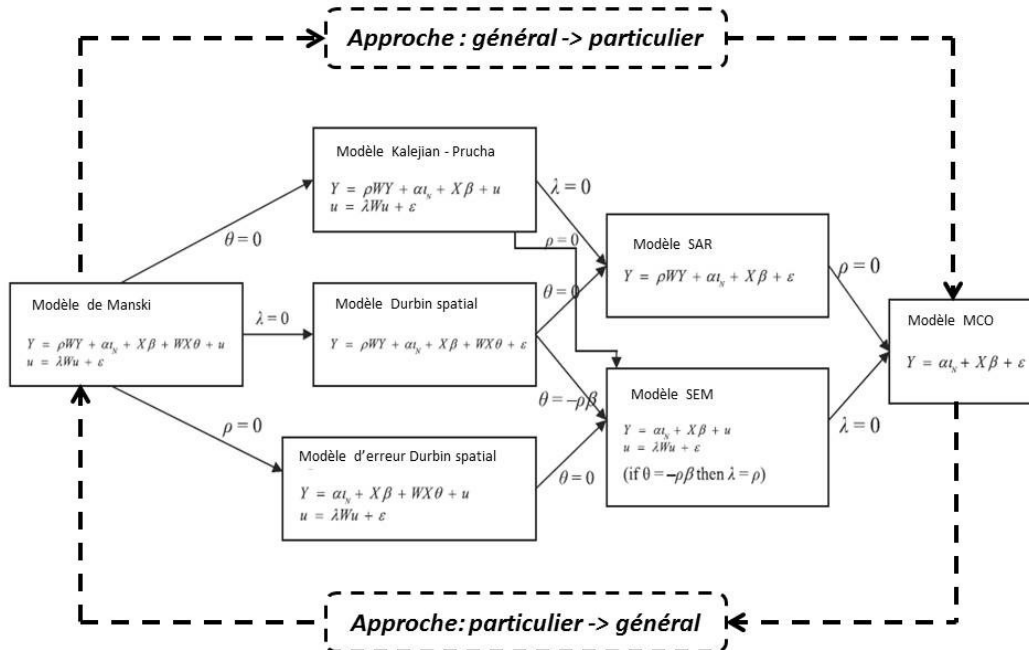
autorégressif mais à sa significativité.¹⁵ Ce test calcule la distance, $(\theta_0 - \theta_1)^2$, entre les estimateurs contraints (θ_1) et non contraints (θ_0). Une trop grande distance entre ces estimateurs implique que l'hypothèse nulle est infirmée et donc rejetée (Dubé et Legros, 2014: 133).

Le test du ratio de vraisemblance (Likelihood Ratio, LR) fonctionne quant à lui selon une logique mixte et permet de comparer la (*log*) vraisemblance d'une spécification contrainte et non contrainte. Une faible distance entre ces deux vraisemblances implique que le test contraint n'est pas différent du test non contraint (Dubé and Legros, 2014a: 132). Si aucune forme de dépendance spatiale n'est identifiée, alors le modèle MCO est conservé (relation B4). Dans le cas contraire, un modèle autorégressif est spécifié (relation B5).

Le nouveau PGD peut également être déterminé en suivant la démarche opposée (du général au particulier)¹⁶. Dans le cas de la seconde démarche (général au particulier, relation A de la figure 3.1), le phénomène observé est d'abord modélisé par un PGD autorégressif (relation A2) intégrant toutes les formes de relations spatiales possibles (modèle de *Manski*). La significativité des paramètres autorégressifs est ensuite testée afin d'en justifier la présence dans le modèle final (relation A3). Les paramètres significatifs seront conservés lors de la modélisation finale, tandis que les autres seront supprimés. La figure 3.2 présente la démarche de sélection du modèle autorégressif afin qu'un nouveau PGD puisse être spécifié (relation A4). Si aucun paramètre autorégressif n'est nécessaire, alors la spécification d'un modèle statique estimé par MCO sera spécifié (relation A5).

¹⁵ Pour plus de détails sur les tests du multiplicateur de Lagrange, test de Wald et le ratio de vraisemblance, voir Dubé et Legros (2014: 131-142).

¹⁶ Pour plus de détails sur l'approche du *général vers particulier*, voir Elhorst (2010 : 13).



adapté de Elhorst, 2010 : 13

Figure 3.2: Démarche de sélection du modèle économétrique

Dans le cadre d'une approche du particulier au général, si les tests statistiques révèlent des traces d'autocorrélation spatiale, plusieurs approches sont possibles. La première mise sur l'insertion d'effets fixes spatiaux, tandis que la seconde propose l'estimation de modèles autorégressifs avec interactions spatiales.

3.1.3 Correction de la dépendance spatiale : les effets fixes et leurs limites

Avant de recourir à des stratégies de modélisation ou des méthodes d'estimation avancées, d'aucuns suggèrent de contrôler la présence de dépendance spatiale ou temporelle par l'inclusion d'effets fixes dans les modèles de régression standards. Ces effets peuvent être utilisés comme variables de contrôle spatiales ou temporelles. On parle alors d'une expansion de Casetti (1972, 1997).

La construction de tels indices permet de mesurer les tendances du marché (dans l'espace et le temps) de manière discrète (par opposition à une approche continue). Les effets fixes spatiaux reposent principalement sur l'agrégation par parcelle d'espace (quartier, région, etc.) qui constituent des choix arbitraires.

Selon Kuminoff (2010), les effets fixes spatiaux représentent la meilleure stratégie de correction des variables omises spatialement corrélées. Ils permettent de scinder un échantillon en sous-groupes et de leur imputer au besoin un terme constant témoignant d'un effet pouvant les affecter selon leur localisation. Cependant, si les effets fixes peuvent contribuer à contrôler la présence d'autocorrélation spatiale, ils sont également accompagnés de certaines limites (Dubé et al., 2011a, Voisin et al., 2010).

D'abord, la fréquence de réalisation des observations (transactions immobilières) peut affecter la possibilité de construire de tels indices (Tu et al., 2004). Ensuite, les effets fixes influencent de façon identique les observations d'un groupe donné, ce qui signifie qu'une possible hétérogénéité du groupe ou la présence d'interinfluences se retrouveront dans le terme d'erreur du modèle de régression. En effet, un seul indicateur pourrait capter un ensemble d'effets pouvant influencer un groupe. Finalement, la délimitation spatiale du groupe n'est pas toujours aisément réalisable (Anselin and Arribas-Bel, 2011). En résumé, l'inclusion d'effets fixes pourrait soulever des problèmes associés à l'agrégation (Anselin, 2002, Anselin and Griffith, 1988, Haining, 2009) et aux aires modifiables (Openshaw,

1983). Les conséquences de ces problèmes se répercutent sur le modèle en affectant la statistique du *goodness of fit* (R^2) de même que les résidus (Tu et al., 2004).

Ainsi, l'utilisation d'effets fixes ne constituerait pas la méthode la plus précise de contrôle de l'autocorrélation spatiale: « The accuracy and precision of the indices will be affected by a number of factors, including the selection of characteristics, the functional form of the hedonic function, behavioral assumptions both on the parameter vectors and the random error terms, and the econometric procedures used to estimate parameters » (Can and Megbolugbe, 1997: 203). Par ailleurs, l'inclusion de tels effets fixes locaux limite sensiblement la considération de la dimension spatiale et donc la portée des analyses régionales. En effet, l'un des attraits principaux de la science régionale réside dans la définition et l'interprétation des effets de débordements spatiaux (*spatial spillovers*) (LeSage, 2014). La prochaine section porte sur l'estimation de modèles autorégressifs destinés à modéliser ces effets.

3.1.4 Correction de la dépendance spatiale : les modèles autorégressifs spatiaux

Les modèles autorégressifs avec interactions spatiales formalisent les liens entre les agents économiques pouvant mener à des comportements collectifs ou des processus d'agrégation qui soulignent le rôle de la localisation (Anselin, 2002). L'intérêt de tels modèles réside dans leur capacité à définir formellement le concept d'effet de débordement spatial, tout en permettant d'en mesurer l'amplitude et d'en tester statistiquement la significativité (LeSage, 2014).

Les nouveaux PGD assurent l'obtention de meilleurs estimateurs en intégrant un effet d'interaction moyen entre les données au PGD. L'estimation d'un modèle de type SAR (*Spatial autoregressive model*) implique l'insertion d'un décalage spatial de la variable

dépendante dans la partie déterministe du modèle et permet de modéliser des effets de débordements (*spillovers*). Ce modèle permet de modéliser les variations de valeurs de la variable dépendante pour une unité spatiale en fonction de la même variable pour les unités spatiales voisines (Elhorst, 2010). Cette structure de données est présentée à l'équation 3.6 sous la forme matricielle (Dubé and Legros, 2014a, LeSage and Pace, 2009b) :

$$\mathbf{y} = \rho \mathbf{S} \mathbf{y} + \alpha \mathbf{1}_n + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (3.6)$$

où les vecteurs \mathbf{y} et $\boldsymbol{\epsilon}$ sont de dimension $[N \times 1]$ tandis que la matrice \mathbf{X} est de dimension $[N \times K]$, où K représente le nombre de variables indépendantes. Le vecteur de paramètres $\boldsymbol{\beta}$ est quant à lui de dimension $[K \times 1]$ (voir Dubé et Legros, 2014 : 123 pour plus de détails). La matrice de pondérations spatiales, notée \mathbf{S} , permet de formaliser les effets de débordements, soit les interdépendances entre les variables dépendantes. L'amplitude de ces effets est représentée par le paramètre ρ . Le modèle SAR est un outil commun dans les applications empiriques spatiales et est habituellement estimé par la méthode du maximum de vraisemblance (ML) ou par la méthode des moments généralisés (GMM). L'un des intérêts du modèle SAR réside dans sa capacité à intégrer les effets de la dépendance spatiale au calcul des effets marginaux d'une variable donnée x_k sur la variable dépendante \mathbf{y} . En effet, dans le cadre d'un modèle de régression standard l'effet marginal est calculé par la dérivée (équation 3.7):

$$\partial y / \partial x_k = \beta_k \quad (3.7)$$

Le modèle SAR permet de calculer un effet total d'une variable explicative donnée sur la variable dépendante. L'effet total représente le cumul des effets de débordements (effet induit) et de l'effet marginal isolé (effet direct). Tel que présenté par LeSage et Pace (2009b), il est obtenu par l'équation 3.8:

$$\partial y / \partial x_k = (\mathbf{I} - \mathbf{S}\rho)^{-1} \mathbf{I}\beta_k \quad 3.8$$

où le terme $(\mathbf{I} - \mathbf{S}\rho)^{-1}$ est habituellement désigné comme le multiplicateur spatial et \mathbf{I} représente une matrice identité de dimension $[N \times N]$ ¹⁷.

La dépendance spatiale peut être modélisée à partir d'une spécification sur le terme de perturbations du modèle de régression MCO. Un modèle de type SEM (*Spatial error model*) renvoie à un décalage spatial de la variable stochastique du modèle suggérant l'omission de variables spatialement corrélées propres aux effets de voisinage, par exemple (Bivand, 1998). Ce modèle permet de capter les caractéristiques environnementales non observées qui encouragent des comportements similaires (Elhorst, 2010). Le PGD du modèle d'erreur spatiale (*Spatial error model* - SEM) est présenté à l'équation 3.9 :

$$\begin{aligned} \mathbf{y} &= \alpha \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda \mathbf{S}\mathbf{u} + \boldsymbol{\epsilon} \end{aligned} \quad (3.9)$$

Les erreurs spatialement dépendantes u sont ainsi en partie fonction des erreurs correspondant aux observations voisines. Cette fois encore une matrice de pondérations spatiales \mathbf{S} permet de formaliser la structure des relations spatiales tandis que le vecteur λ renseigne sur leur force.

Lorsque estimé par la méthode des moindres carrés ordinaires, le modèle SEM présente des coefficients β_k sans biais mais de variance inefficace affectant les tests statistiques. Ainsi, comme le modèle SAR, l'estimation du modèle SEM se fait principalement par la méthode du maximum de vraisemblance (ML) ou par la méthode des

¹⁷ Pour plus de détails sur le calcul des effets directs, indirects et totaux le lecteur intéressé peut se référer à Dubé et Legros (2014 : 125).

moments généralisés (GMM). En revanche, le calcul de l'effet marginal est similaire à celui du modèle standard (équation 3.7).

Jusqu'ici, seule la dimension spatiale a été abordée. La prochaine section se penche sur les défis associés à l'inclusion simultanée de la dimension temporelle

3.2 MODÉLISATION DES PROCESSUS SPATIOTEMPORELS

3.2.1 Conjuguer temps et espace

Malgré la complexité de la tâche, l'inclusion de la dimension spatiale à l'analyse immobilière devient commune depuis les travaux de Can (1992). En revanche, l'inclusion simultanée de la dimension temporelle fait défaut. En appliquant de manière quasi-mécanique les méthodes économétriques spatiales aux données en coupes transversales empilées, donc en omettant les phénomènes de dépendance temporelle, le chercheur postule des relations de proximité spatiale mais ignore que ces effets sont également situés dans le temps. Par le fait même, des relations spatiales entre des observations temporellement distantes sont postulées. Une telle hypothèse suppose l'existence de relations spatiales abusives et représente une violation de la logique temporelle (*arrow of time*) selon laquelle les observations passées influencent les observations futures, alors que l'inverse est contre nature (Thanos et al., 2016). En outrepassant cette logique temporelle, des relations spatiales abusives sont postulées au risque de nuire à la définition des processus de causalité et donc au terme de perturbations du modèle de RLM. Plusieurs options sont possibles pour considérer simultanément espace et temps.

Sur le même principe que pour le contrôle de la dimension spatiale, l'inclusion d'effets fixes temporels permet de considérer certains facteurs tels que l'inflation (effet

temporel) ou encore le phénomène d'obsolescence associé à l'âge du logement (Huang et al., 2010). Comme pour le cas des effets fixes spatiaux, la définition de ces variables constitue des choix arbitraires et est soumise à des limitations similaires, un problème d'agrégation notamment.

La définition d'une matrice de pondérations spatiotemporelles est généralement proposée. La difficulté réside dans la structuration des données spatiotemporelles collectées de manière non répétée et particulièrement dans la conjugaison des dépendances spatiale et temporelle. Néanmoins, la non-répétition des observations rend difficile l'intégration conjointe des dimensions spatiales et temporelles lors de la modélisation. Ainsi, la démarche d'intégration des relations spatiotemporelles fait débat (Dubé et al., 2013a, Dubé and Legros, 2013b, Dubé and Legros, 2013a, Smith and Wu, 2009). Le défi réside dans l'exploitation de microdonnées sans recourir à l'agrégation spatiale afin d'exploiter l'information complète fournie par chaque observation (Thanos et al., 2016). Pour ce faire, la matrice de pondérations spatiales doit être ajustée afin de respecter la logique unidirectionnelle temporelle.

Les premières tentatives imposent une restriction temporelle aux relations spatiales. Les relations de dépendance avec les observations futures peuvent être postulées comme étant nulles et avec les relations passées limitées à une période donnée. Cette stratégie est notamment proposée par Can et Megbolugbe (1997) ainsi que Brandt et Maennig (2011).

Pace et al. (1998, 2000) proposent pour leur part la spécification d'un modèle destiné à filtrer les effets spatiaux et temporels des observations voisines afin de produire des estimateurs fiables (modèle STAR).

Les auteurs proposent de construire des filtres basés sur la construction de matrices spatiales (\mathbf{S}), temporelles (\mathbf{T}) et spatiotemporelles (\mathbf{TS} et \mathbf{ST})¹⁸. L'idée générale est donc, pour la valeur d'une observation donnée, de soustraire la part de variation de sa valeur générée par les observations voisines précédentes : « to take the current value of a variable at a location and subtract an average of past, surrounding values scaled by a constant less than 1 » (Pace et al., 1998: 17). Le modèle STAR vise principalement à réduire le nombre de variables spatiales et temporelles de contrôle (effets fixes) incluses dans les modèles de régression et susceptibles d'influencer les estimations. Néanmoins, cette approche laisse certaines ambiguïtés sur la prise en compte des processus spatiaux et temporels et leur interprétation en raison notamment du nombre de matrices considéré.

D'autres propositions visent donc à spécifier une matrice de pondérations spatiotemporelles générale intégrant simultanément les contraintes spatiales et temporelles.

3.2.2 La matrice de pondérations spatiotemporelles

La spécification d'une matrice spatiotemporelle unique implique tout de même de distinguer la nature différente des relations de dépendance spatiales : multidirectionnelles (simultanées) et unidirectionnelles (temporelles). Cette matrice de pondérations spatiotemporelles est alors issue de deux matrices distinctes : i) une matrice de pondérations spatiales et ii) une matrice de pondérations temporelles.

¹⁸ Il est à noter que pour faciliter les calculs, les matrices \mathbf{S} et \mathbf{T} sont triangulaires. Les observations y sont organisées chronologiquement de sorte que l'observation la plus ancienne se trouve sur la première ligne. Ainsi, les éléments situés au-dessus de la diagonale de la matrice représentent le futur et sont donc nuls. Le produit des matrices \mathbf{S} et \mathbf{T} crée une matrice \mathbf{ST} également triangulaire. Il en va de même pour la matrice \mathbf{TS} .

Comme pour la matrice de pondérations spatiales, la matrice de pondérations temporelles repose également sur la définition d'une proximité des observations, mais dans le temps cette fois. Une distance seuil peut être imposée par le modélisateur afin de ne considérer que les observations situées dans un rayon temporel défini. La proximité temporelle permet de définir jusqu'à quel point des observations passées peuvent influencer des observations futures. L'équation 3.10 présente une démarche de calcul des éléments de la matrice \mathbf{T} proposée par Dubé et Legros (2013a).

$$t_{ij} = \begin{cases} |v_i - v_j|^{-\gamma} & \text{si } |v_i - v_j| < \bar{v} \quad \forall i \neq j \\ 1 & \text{si } v_i = v_j \quad \forall i \neq j \\ 0 & \text{sinon} \end{cases} \quad (3.10)$$

Selon le même principe que pour la matrice \mathbf{S} , \bar{v} représente la distance temporelle seuil et l'inverse de la période de temps durant laquelle la dépendance effective est utilisée. Finalement, v_i et v_j sont les moments durant lesquels les observations i et j sont réalisées. En classant chronologiquement les observations, la première ligne de la matrice temporelle (et aussi spatiale) correspondra à la plus ancienne observation collectée. Le triangle supérieur à la diagonale représentera ainsi les liens entre une observation et les observations collectées plus tard (le futur). Le triangle supérieur de la matrice peut être composé de 0 lorsque le chercheur postule que ces relations sont impossibles (équation 3.11). Si un effet d'anticipation est postulé, les éléments de la diagonale supérieure seront de valeur non nulle selon un seuil temporel donné (voir notamment Thanos et al., 2016).

$$\mathbf{T} = \begin{bmatrix} 1 & t_{12} & \cdots & t_{1i} & \cdots & t_{1n} \\ t_{21} & 1 & \cdots & t_{2i} & \cdots & t_{2n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{i1} & t_{i2} & \cdots & t_{ij} & \cdots & t_{jn} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nj} & \cdots & 1 \end{bmatrix} \quad (3.11)$$

D'autres, tels que Dubé et Legros (2013b, 2013a) ou encore Smith et Wu (2009) et Huang et al. (2010) avant eux, proposent de traiter les données de transactions immobilières en coupes transversales empilées dans le temps. Ceux-ci proposent et justifient la construction d'une matrice agrégée combinant les deux types de matrices (effet du temps et de l'espace). La matrice spatiotemporelle (\mathbf{W}) est construite par le produit de Hadamard, donc par le produit des matrices spatiale (\mathbf{S}) et temporelle (\mathbf{T}) (équations 3.12 et 3.13):

$$\mathbf{W}_{(N_T \times N_T)} = \mathbf{S}_{(N_T \times N_T)} \odot \mathbf{T}_{(N_T \times N_T)} \quad (3.12)$$

$$\mathbf{W}_{(N_T \times N_T)} = \begin{bmatrix} 0 & s_{12} \times t_{12} & s_{13} \times t_{13} & \cdots & s_{1N_T} \times t_{1N_T} \\ s_{21} \times t_{21} & 0 & s_{23} \times t_{23} & \cdots & s_{2N_T} \times t_{2N_T} \\ s_{31} \times t_{31} & s_{32} \times t_{32} & 0 & \cdots & s_{3N_T} \times t_{3N_T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{N_T1} \times t_{N_T1} & s_{N_T1} \times t_{N_T2} & s_{N_T3} \times t_{N_T3} & \cdots & 0 \end{bmatrix} \quad (3.13)$$

Dubé et Legros (2013b, 2013a) soulignent le rôle de cette spécification pour calculer des statistiques d'autocorrélation spatiales et les tests d'hypothèses en plus de souligner que la période de temps influence ces paramètres. Ils notent également en analysant des données immobilières pour la ville de Québec (1986-1996) que l'impact de cette spécification peut influencer les coefficients estimés de même que leur variance et peut donc avoir un impact sur leur significativité. Finalement, ils observent ces résultats quels que soient les critères de relations spatiales (contiguïté ou distance) ou de temps (seuil de temps considéré) : « Even if spatial autocorrelation remains significant, the results suggest that time considerations may influence the magnitude of the estimated phenomena in several ways » (Dubé and Legros, 2013b: 29). En comparant l'impact d'une première matrice exclusivement spatiale, \mathbf{S} , et d'une seconde spatiotemporelle, \mathbf{W} , sur la statistique

du I de Moran, Dubé et Legros constatent une variation dans le coefficient estimé ainsi que sa variance, ce qui implique un impact sur la significativité du test statistique : « Considering the spatial dimension alone when data have an important temporal dimension can lead to the overestimation of spatial dependence in residuals and, consequently, can lead to overcorrection using the appropriate spatial econometric approaches » (Dubé and Legros, 2013b: 26).

La formulation d'une matrice de pondérations spatiotemporelles permet de respecter le principe multidirectionnel de la dimension spatiale tout en respectant la logique unidirectionnelle du temps. L'approche permet donc d'interpréter les phénomènes régionaux et leur structure de dépendance spatiale tout en considérant leur évolution dans le temps. Tel que supposé plus tôt, omettre la dimension spatiale pourrait être synonyme de relations de dépendance abusives entre des observations temporellement éloignées et encourager un problème de sur-connectivité (Smith and Wu, 2009). Dans une telle situation, la détection de l'autocorrélation spatiale pourrait être biaisée, de même que sa correction. Des répercussions sur les modèles économétriques spatiaux pourraient en résulter.

Dans le cas d'un modèle SAR (variable dépendante spatialement décalée), l'estimation du coefficient ρ faussé a des répercussions sur les autres coefficients estimés. Ainsi, le calcul de l'effet total d'une variable donnée sur la variable dépendante serait affecté et mener ainsi à des conclusions d'études erronées. Dans le cas d'un modèle SEM (terme d'erreurs spatialement décalé), l'effet du coefficient λ faussé se répercute sur la précision de la variance des estimateurs. En conséquence, la significativité des variables d'intérêt est également faussée. Ultiment, la mauvaise interprétation des relations de dépendance spatiotemporelles possède des effets directs sur les résultats de recherche. Dans le cadre de l'analyse régionale ou territoriale, il pourrait par exemple en résulter une évaluation contre-productive de politiques publiques.

Au final, bon nombre de travaux ignorent encore la dimension temporelle dont l'impact demeure incertain sur les estimations. Certaines études de cas unique ont été mentionnées, mais leurs résultats ne peuvent être généralisés. La méta-analyse pourrait permettre de dépasser certaines limites des travaux précédents. Cette stratégie de recherche est présentée dans la section suivante.

3.3 MÉTA-ANALYSE ET MÉTA-RÉGRESSION

Afin d'évaluer l'effet de l'omission du temps sur les statistiques spatiales, la recherche propose la réalisation d'une méta-analyse. Cette stratégie vise à expliquer la variation des statistiques spatiales calculées ou estimées en fonction d'un ensemble de facteurs associés au PGD. Les rudiments de la méta-analyse sont abordés dans la prochaine section.

3.3.1 Définition

Le projet repose en partie sur la réalisation d'une méta-analyse (MA) basée sur les résultats d'études scientifiques associées aux valeurs immobilières. L'objectif est de déterminer quels sont, parmi les paramètres des études ainsi que leur PGD, les éléments qui influencent les statistiques spatiales afin de répondre aux différentes questions spécifiques de recherche proposées. La MA représente une approche pertinente d'analyse puisqu'elle permet de considérer et d'expliquer un grand nombre de statistiques spatiales issues d'études de cas empiriques.

La MA se révèle une méthode adéquate dans un contexte d'abondance de connaissances scientifiques dans le but de réaliser une synthèse exhaustive des résultats,

mais également à des fins de transfert de résultats issus d'études de cas variés observant des phénomènes similaires (Florax et al., 2002). Cette démarche se définit donc comme une analyse statistique permettant des résultats d'études qui, dans le cas de l'économie, estiment une élasticité particulière (Schipper et al., 2002). Plus précisément, la MA est un outil quantitatif utilisé dans le but d'expliquer de manière économétrique la variation dans les estimations empiriques d'un champ de recherche particulier (Ghinamo, 2012). Le terme MA est introduit par Glass (1976) qui la définit ainsi :

Meta-analysis refers to the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the causal, narrative discussions of research studies which typify our attempt to make sense of the rapidly expanding research literature. (Glass, 1976 dans Florax et al., 2002 : 27).

Bal et Nijkamp (2001) lui attribuent une définition plus vaste en proposant l'application de méthodes quantitatives mais également qualitatives. Ils soulignent également la nécessité d'appliquer cette démarche à un ensemble d'études clairement défini. La MA est donc :

A scientific investigation of a well-defined set of previously published individual studies concerning a certain subject, with the aim to apply qualitative as well as quantitative review techniques in order to make an integral cross-sectional or comparative study of the available material, leading to more insights on the chosen subject of research and ending up with a systematic synthesis (Bal and Nijkamp, 2001: 369).

La MA permet également la création de connaissances synthétiques appropriées au transfert de valeur (value transfer).¹⁹ Ce transfert de valeur vise l'atteinte de nouveaux objectifs de recherche basés sur l'accumulation de connaissances synthétisées par la MA et permet donc de transférer les valeurs des paramètres d'une étude de cas à une autre sous

¹⁹ Le terme *value transfer* est ici traduit par transfert de valeur et fait référence aux valeurs de paramètres des études de cas utilisées.

certaines conditions. Le transfert direct de valeur n'est pas l'objectif premier de la présente recherche, néanmoins ses conclusions dépasseront le champ de l'analyse des prix hédoniques des valeurs immobilières et pourront s'appliquer à toutes formes d'analyses spatiales utilisant des données en coupes transversales empilées.

3.3.2 Méta-régression et effet de taille (Effect Size)

La MA n'est pas une méthode de recherche statistique à proprement parler. En revanche, elle repose généralement sur le calcul d'un effet moyen et sur le principe de la régression linéaire multiple aussi nommée méta-régression (MR) dans le contexte de la MA, qui elle est la méthode de recherche.

La MA vise généralement à estimer un effet de taille (*Effect Size, ET*) pour un ensemble d'études empiriques afin d'expliquer la variation observée des effets entre celles-ci. L'ET (équation 3.14) correspond à l'effet moyen pondéré par l'inverse de sa variance (Borenstein et al., 2009) :

$$ET = \frac{\sum_{z=1}^Z P_z \pi_z}{\sum_{z=1}^Z P_z} \quad (3.14)$$

où, π_z représente l'effet pour une étude empirique donnée z et P_z représente le poids défini par l'inverse de la variance : $1/v_{\pi_z}$.

Au-delà du calcul d'un effet moyen, la MA vise également à expliquer la fluctuation de cet effet entre différentes études (ou échantillons) par le biais d'un modèle de méta-régression MR (Harbord and Higgins, 2008b). Contrairement à un modèle de régression classique permettant d'estimer la part de variation d'une variable expliquée par le modèle

(statistique R^2), le modèle de MR permet d'extraire et d'expliquer la proportion de variation inter-études, ψ^2 , de la variation totale d'une variable donnée (Higgins and Thompson, 2002b).²⁰ D'après Borenstein et al. (2009) et tel qu'illustré à la figure 3.3²¹, la statistique ψ^2 peut être comprise comme une mesure d'inconsistance entre les résultats d'études.

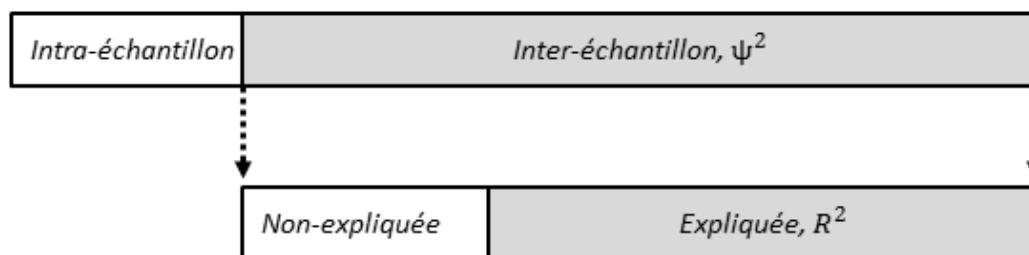


Figure 3.3: Variation expliquée par la méta-régression

La MR est utilisée afin d'expliquer la variation du coefficient d'une étude donnée, z , dans un ensemble de m recherches. Ce coefficient devient la variable dépendante du modèle de MR. Ainsi, dans un ensemble de m études s'appropriant le modèle de prix, le coefficient donné γ devient la variable à expliquer en fonction de différentes caractéristiques de l'étude. Selon le principe de base de la régression dans une MA, l'écart-type de la variable expliquée, σ_i , est connu. Partant de cela, il existe plusieurs modèles de MA (voir Harbord and Higgins, 2008b).

²⁰ L'auteur réfère à la variance inter-études avec le paramètre I^2 . Il est ici renommé ψ^2 pour le distinguer de l'indice I de Moran.

²¹ Adapté de Borenstein et al. (2009: 202)

Le premier correspond à une MA avec effets fixes, μ , de sorte que $\gamma_i = \mu + v_i$ où, $v_i \sim N(0, \sigma_i^2)$.

Le second modèle est un modèle à effets aléatoires qui autorise la taille de l'effet, μ_i , à varier entre les études en assumant qu'ils ont une distribution normale autour d'une moyenne. Ainsi, $\gamma_i = \mu + u_i + v_i$, où $u_i \sim N(0, \tau^2)$ et $v_i \sim N(0, \sigma_i^2)$. Ici, τ^2 représente la variance entre les études et doit être estimé depuis les données²².

Le troisième modèle est une extension du modèle à effets fixes dans lequel l'effet moyen, μ est remplacé par un régresseur linéaire, $x_i\delta$. Le modèle est alors formalisé par $\gamma_i = x_i\delta + v_i$ où cette fois encore ε_i est normalement distribué et de variance σ_i^2 .

Finalement, une extension du modèle à effets aléatoires est également possible autorisant la présence d'hétérogénéité dans les résidus (dans la variance entre les études non expliquée par des covariables). Le modèle de MR est alors de la forme $\gamma_i = x_i\delta + u_i + v_i$ où $u_i \sim N(0, \tau^2)$ et $v_i \sim N(0, \sigma_i^2)$.

3.3.3 Avantages et limites de la méta-analyse

La MA possède des avantages et des limites. Elle permet notamment l'analyse d'un champ de recherche au-delà des incohérences possibles entre différentes études de cas. Généralement, le recours à l'étude de cas ne vise pas à confirmer des connaissances existantes mais à en produire de nouvelles à travers la répétition d'expériences. Néanmoins, dans les sciences sociales, la réalisation d'expériences en milieu contrôlé est habituellement

²² Le lecteur intéressé pourra se référer à Harbord et Higgins (2008 : 92-93) qui décrivent en détail la méthode d'estimation de la variance inter-études, τ^2 , par la méthode des moments, de la maximisation de la log-vraisemblance des résidus (MLVR) et un estimateur bayésien empirique.

irréalisable, ce qui mène à des problèmes d'incohérence entre les résultats de recherche. En effet, les efforts scientifiques réalisés dans le cadre d'études de cas visent à ce que la recherche empirique permette de générer de nouvelles connaissances afin de créer ou contribuer cumulativement au développement de théories exploratoires, ou incrémentalement aux théories existantes, sur une problématique donnée (Bal and Nijkamp, 2001).

Plusieurs raisons expliquent les incohérences entre les études : 1) une mauvaise manipulation de l'information cumulée ; 2) une collecte de données ou un processus d'observation défaillants ; 3) une mauvaise application des méthodes d'analyse et 4) un cadre méthodologique inapproprié. Dans la perspective d'une étude quantitative en sciences sociales, il faut ajouter à ces erreurs les problèmes associés à un mauvais échantillonnage, aux erreurs de mesure et à l'étendue de la variance des résultats (Bal and Nijkamp, 2001). Pour pallier ce dernier problème, Hunter et Schmidt (1990 dans Bal et Nijkamp, 2001) recommandent dans le cas d'une étude singulière de recourir aux intervalles de confiance et dans le cas d'une revue de plusieurs études d'utiliser la MA.

La MA permet de surmonter des problèmes tels que le manque de grandes bases de données permettant de tendre vers la généralisation des résultats, ou encore le problème de l'incertitude de l'information et de valeur des données. C'est à ces fins qu'historiquement la MA a été introduite dans les sciences naturelles et la médecine, puis dans les sciences sociales dans les années 70. De plus, elle est utile en termes de gain d'efficacité des recherches empiriques en permettant de transférer, même partiellement, les résultats pratiques tout en réduisant grandement les coûts associés à la réalisation d'une étude de grande envergure puisqu'elle supprime la collecte de données sur le terrain et analyse plutôt les résultats d'études de cas variées. De plus, sans nécessairement prétendre à la généralisation des résultats, elle permet de faire un pas intéressant vers celle-ci en autorisant une comparaison d'études de cas aussi variées que diversifiées. En effet, par

principe la MA ne tente pas de mettre en relation des éléments contextuels propres à chaque cas, mais vise plutôt à expliquer les coefficients des modèles statistiques estimés. Au final, il devient possible d'expliquer les différences entre des résultats normalement non généralisables. Par ailleurs, la MA permet d'intégrer aux études réalisées des éléments contextuels qui les caractérisent. Il est ainsi possible de déterminer et d'expliquer si ces caractéristiques exercent une influence sur les résultats observés.

Ghinamo (2012) résume les principales limites de cette approche en quatre points. La MA souffre d'un problème de disparités dans les caractéristiques d'échantillonnage, les méthodes d'estimation et de structuration des recherches, et laisse supposer une incompatibilité des résultats de recherche regroupés par la MA. Cependant, l'auteur précise que tant que les valeurs estimées utilisées dans la MA sont considérées selon une mesure sans unité (un coefficient par exemple) et que les structures de recherche sont contrôlées dans le modèle de régression de la MA, alors le problème d'hétérogénéité n'existe plus, ce qui peut finalement être considéré comme un avantage sur une revue de la littérature classique. Ensuite, la MA pose un problème de biais d'échantillonnage. Le chercheur doit porter une attention particulière aux études incluses dans son analyse et ne considérer que les recherches pertinentes. Cette fois encore, ce biais étant clairement spécifié par le chercheur lors de la construction de l'échantillon, l'omission ou l'inclusion de certaines études est ainsi justifié, ce qui permet d'afficher sans équivoque les biais potentiels. Les biais de publication sont en revanche plus délicats à appréhender. Ces biais peuvent surévaluer, ou sous-évaluer, certaines estimations dans la littérature. Cependant, la MA permet de tester²³ et contrôler ces biais en considérant notamment les revues publiant les

²³ Ghinamo (2012) présente le *Precision Effect Test* pour tester la présence de biais de publication. Ce test constitue une régression des estimateurs d'élasticité sur leur écart-type. Un test d'hypothèse (*t*-test) peut ensuite être réalisé sur le nouveau coefficient estimé. Suite à ce test, si le nouveau coefficient est égal à 0 alors il n'y a pas de biais.

études dans le modèle d'analyse. Finalement, il est difficile d'évaluer l'impact des études de moindre qualité dans l'échantillon utilisé pour la MA. Pour contourner ce problème, le modèle d'analyse doit contrôler les méthodes utilisées (type de modèle économétrique et méthode d'estimation). Une fois ces variables incluses dans le modèle, la MA permet de tester l'influence des méthodes de recherche sur les résultats observés dans la littérature. Finalement, sans être parfaite, la MA se révèle une approche intéressante puisque bon nombre de ses limites tournent à son avantage.

CHAPITRE 4

REVUE DE LA LITTÉRATURE

Les notions de distance et de proximité constituent des éléments centraux dans l'analyse des dynamiques régionales ou urbaines. En parcourant le chapitre précédent, le lecteur aura pu se familiariser avec les démarches de modélisation spatiale et spatiotemporelle destinées à appréhender ces phénomènes.

La matrice de pondérations spatiales permet la formalisation des relations spatiales et représente par le fait même la pierre angulaire de l'économétrie spatiale. La spécification de cette matrice est faite de manière exogène et influence la compréhension d'un phénomène observé. Elle est pourtant utilisée tout au long de la démarche de modélisation. Elle est en effet une partie intégrante du calcul de l'indice I de Moran destiné à mesurer et tester statistiquement la présence d'autocorrélation spatiale, ainsi que dans la formulation de modèles autorégressifs destinés à en corriger les effets.

Pourtant, les problématiques empiriques régionales et urbaines ne se limitent pas à la seule dimension spatiale. Le temps a été identifié comme un élément central de la définition des mécanismes de causalité. En d'autres termes, les phénomènes spatiaux sont situés dans le temps et sont fonction de leur état précédent. Ainsi, les relations de dépendance temporelles se distinguent par leur nature unidirectionnelle. L'un des défis réside donc dans la conjugaison de ces deux dimensions afin de garantir une modélisation réaliste d'un phénomène régional particulier.

Les prochaines lignes visent à présenter une brève revue de la littérature de travaux ayant recours à l'analyse hédonique pour expliquer les valeurs foncières. Une première section se concentrera sur les analyses empiriques strictement spatiales. Étant donné la volumineuse littérature sur le sujet, cette section ne prétend pas être exhaustive. Malgré les problèmes désormais connus causés par les dépendances spatiales et temporelles, elle permet néanmoins de souligner que les études ignorant le temps demeurent fréquentes. Dans un second temps, certains travaux misant sur une approche spatiotemporelle sont présentés.

4.1 ÉTUDES UTILISANT UNE APPROCHE EN COUPES TRANSVERSALES

La formalisation des relations spatiotemporelles est une étape importante. Les conséquences d'une mauvaise formalisation des phénomènes spatiaux sur les estimateurs et leur variance ont été présentées plus tôt. Pourtant, une partie non négligeable des études empiriques ignore la dimension temporelle dans leur stratégie de modélisation. Au mieux, le temps est considéré à travers la construction de variables binaires identifiant la période de transaction. Le fait est que les analyses hédoniques traitent une grande variété de problématiques et constituent un outil commun d'analyse de l'impact de politiques publiques influençant les trajectoires de développement. À ce titre, elles sont utilisées afin de mettre en évidence l'effet de certaines externalités liées à des stratégies de développement urbain. Dans une perspective plus large, elles permettent aussi de comprendre et d'interpréter économiquement l'impact de caractéristiques propres à un milieu de vie. Le tableau 4.1 recense certaines études empiriques utilisant une approche strictement spatiale.

Tableau 4.1: Articles considérant l'analyse en coupe transversale (omission du temps)

Auteurs	Marché	Secteur de l'étude	Étendue temporelle	
Andersson et al. (2010)	Résidentiel	Lerum (Suède)	1996	2006
Anselin et Lazano-Gracia (2008)	Résidentiel	Californie (US)	1999	1999
Armstrong et Rodriguez (2006)	Résidentiel	Massachusetts (US)	1992	1993
Bell et Bockstael (2000)	Résidentiel	Nothern Anne Arundel County (US)	1993	1996
Bin et al. (2008)	Résidentiel	New-Hanover county (US)	1995	2002
Bowen et al. (2001)	Résidentiel	Cuyahoga county (US)	1 an ²⁴	
Boxall et al. (2005)	Résidentiel	Calgary (Canada)	1994	2001
Brandt and Meannig (2011)	Résidentiel	Hamburg (Allemagne)	2002	2008
Brandt and Meannig (2012)	Résidentiel	Hamburg (Allemagne)	2002	2008
Brasington (2004)	Résidentiel	Ohio (US)	1991	1991
Brasington et Samara (2008)	Résidentiel	Ohio (US)	2000	2000
Can (1990)	Résidentiel	Columbus (US)	1980	1980
Capozza et al. (2005)	Résidentiel	50 États différents (US)	1988	1999
Carrion-Flores et Irwin (2010)	Résidentiel	Calvert County (US)	1997	2001
Cavaillès et al. (2009)	Résidentiel	Dijon (France)	1995	2002
Cohen et Coughlin (2008)	Résidentiel	Atlanta (US)	2003	2003
Conway et al. (2010)	Résidentiel	Los Angeles (US)	1999	2000
Crespo et Grêt-Regamey (2013)	Résidentiel	Zurich (Suisse)	2004	2005
Dekkers et van der Straaten (2009)	Résidentiel	Amsterdam (Hollande)	1999	2003
Dermisi et McDonald (2010)	Bureaux	Chicago (US)	1996	2007
Dermisi et McDonald (2011)	Bureaux	Chicago (US)	1996	2008
Des Rosiers et al. (2010)	Résidentiel	Québec (Canada)	1993	1997
Donovan et al. (2007)	Résidentiel	Colorado Springs (US)	1999	2002
Dubé et al. (2012)	Résidentiel	Québec (Canada)	1986	1996
Fernandez-Alviles et al. (2012)	Résidentiel	Madrid (Espagne)	2009	2009
Geoghegan et al. (2003)	Résidentiel	3 comtés du Maryland (US)	1993	1996
Ham et al. (2013)	Résidentiel	El Paso County (US)	2005	2007
Hansz et Hayunga (2012)	Résidentiel	Pinehurst (US)	2002	2004
Helbich et al. (2014)	Résidentiel	Autriche	1998	2009
Hindley et al. (2013)	Résidentiel	Pinellas county (US)	2000	2006
Holmes et al. (2006)	Résidentiel	Sparta (US)	1992	2002
Hui et al. (2007)	Résidentiel	Pénins. de Kowloon (Hong Kong)	2000	2001
Kestens et al. (2006)	Résidentiel	Québec (Canada)	1993	2001
Kim et al. (2003)	Résidentiel	Seoul (Corée)	1993	1993
Kim et Goldsmith (2009)	Résidentiel	Craven County (US)	1990	2003
Kostov (2009)	Sol	Irlande du Nord	1996	1999
Kryvobokov (2010)	Résidentiel	Lyon (France)	1997	2008
Kuethé (2012)	Résidentiel	Milwaukee (US)	2005	2005
Landry (2011)	Résidentiel	Savannah (US)	1990	1999

²⁴ L'année de l'étude n'est pas précisée. Les modèles incluent trois variables d'effets fixes trimestriels. Avec le trimestre de référence, la base de données couvre donc une période de quatre trimestres, soit un an.

Auteurs	Marché	Secteur de l'étude	Étendue temporelle	
Lazrak et al. (2014)	Résidentiel	Zaanstad. (Allemagne)	1985	2007
Leguizamon et Ross (2012)	Résidentiel	Ohio (US)	2000	2000
Li et Saphores (2012)	Résidentiel	Los Angeles (US)	2003	2004
Liao et Wang (2012)	Résidentiel	Changsha (Chine)	2009	2008
Löchl et al. (2010)	Résidentiel	Zurich (Suisse)	2004	2005
Mihaescu et Vom Hofe (2012)	Résidentiel	Cincinnati (USA)	2009	
Militino et al. (2004)	Résidentiel	Pampelune (Espagne)	2000	2000
Neelawala et al. (2013)	Résidentiel	Mount Isa (Australie)	2003	2010
Nelson (2010)	Résidentiel	Maryland (US)	2000	2000
Newburn et al. (2006)	Sol	Sonoma county (US)	2000	2000
Noonan (2007)	Résidentiel	Chicago (US)	1990	1999
Osland et al. (2007)	Résidentiel	Rogaland (Norvège)	1997	2001
Osland et Pryce (2012)	Résidentiel	Glasgow (Écosse)	2007	2007
Osland et Thorsen (2013)	Résidentiel	Rogaland (Norvège)	1997	2001
Osland, I. (2010)	Résidentiel	Norvège	1997	2002
Parent et Vom Hofe (2013)	Résidentiel	Miami (US)	2005	2005
Paterson et Boyle (2002)	Résidentiel	Simsbury et Avon (US)	1997	1999
Patton et McErlean (2003)	Sol	Irlande	1996	1999
Perdomo (2011)	Résidentiel	Bogota (Colombie)	2008	2008
Price et al. (2010)	Résidentiel	Grand County (US)	1995	2006
Salvi (2003)	Résidentiel	Zurich (Suisse)	1995	2007
Samarasinghe et Sharp (2010)	Résidentiel	North Shore (NZ)	2006	2006
Sander et Haight (2012)	Résidentiel	Dakota County (US)	2005	2005
Saphores et Li (2012)	Résidentiel	Los Angeles (US)	2003	2004
Sedgley et al. (2008)	Résidentiel	Howard County (US)	2002	2002
Seya et Tsutsumi (2013)	Résidentiel	Japon	2003	2010
Tsutsumi et Seya (2009)	Résidentiel	Japon	1999	1999
Walker (2016)	Commercial/résidentiel	Shelby County (US)	2010	
Wallner (2012)	Résidentiel	Sydney (Australie)	2008	2008
Walsh et al. (2011)	Résidentiel	Washington et Californie (US)	1996	2004
Wilhelmsson (2002)	Résidentiel	Stockholm (Suède)	2000	2001
Yu et al. (2012)	Résidentiel	Knox County (US)	1996	2005
Zhu et al. (2011)	Résidentiel	Lucas County (US)	1998	1998

À titre d'exemple, les nuisances sonores sont largement considérées dans la littérature. Le bruit lié à l'activité aérienne (avion et aéroport) génère habituellement un impact négatif (Cohen and Coughlin, 2008, Dekkers and van der Straaten, 2009) ou modéré (Salvi, 2003) sur les valeurs résidentielles. Des conclusions similaires sont avancées par Armstrong et Rodriguez (2006) et Walker (2016) pour la proximité aux voies de chemin de fer. Néanmoins, les conséquences de ces nuisances ne seraient pas constantes dans l'espace

(Brandt and Maennig, 2011) ou selon le mode de transport (Andersson et al., 2010). D'ailleurs, la fréquence de passage des transports en commun affecte négativement également la prime associée à leur proximité des résidences (Des Rosiers et al., 2010, Perdomo, 2011).

La description de l'environnement et de son impact sur les valeurs résidentielles fait aussi l'objet d'une large considération comme en témoignent les études qui ciblent la proximité d'antennes cellulaires (Brandt and Maennig, 2012b), de friches industrielles (Mihaescu and vom Hofe, 2012), la vue et le paysage (Cavailhès et al., 2009, Hindsley et al., 2013, Paterson and Boyle, 2002, Wallner, 2012) ou la couverture végétale (Conway et al., 2010, Li and Saphores, 2012, Saphores and Li, 2012). La perception des acheteurs peut également être reflétée à travers l'impact de risques naturels aussi variés que les inondations (Bin et al., 2008, Samarasinghe and Sharp, 2010), les incendies (Donovan et al., 2007) ou encore les invasions d'insectes (Holmes et al., 2006, Price et al., 2010). La pollution fait également l'objet d'une attention particulière (Kestens et al., 2004, Anselin and Lozano-Gracia, 2008, Fernández-Avilés et al., 2012, Kim et al., 2003, Neelawala et al., 2013).

Finalement, une littérature abondante utilise le modèle de prix hédoniques et différentes thématiques de recherche comme prétexte pour des considérations méthodologiques ou l'amélioration des démarches de modélisation. À ce titre, Ham et al. (2013) soulignent l'importance de ne pas postuler l'espace comme homogène sous peine de nuire aux paramètres estimés et de biaiser les résultats de recherche. Helbich et al. (2014) prétendent en revanche que celle-ci est trop complexe pour être modélisée par des indicateurs régionaux ou locaux tandis que Dubé et al. (2012) proposent la méthode d'expansion de Fourier sur les coordonnées géographiques afin de contrôler les effets de la présence d'autocorrélation spatiale. D'autres suggèrent la régression par quantile pour explorer les variations des prix implicites selon la distribution des valeurs résidentielles

(Liao and Wang, 2012, Kostov, 2009, Amédée-Manesme et al., 2016). La littérature ne néglige pas de mentionner la nécessité de contrôler les effets de la dépendance spatiale (Bowen et al., 2001, Liu, 2013, Zhu et al., 2011) ou encore d'identifier justement les processus spatiaux en action (Parent and vom Hofe, 2013, Wilhelmsson, 2002).

Les effets spatiaux y sont intégrés par l'inclusion d'effets fixes locaux, ainsi que par l'inclusion de variables spatiales mesurant la distance ou la proximité à des aménités. Les effets fixes spatiaux réfèrent à des niveaux d'agrégation différents. À titre d'exemple, Dekkers et van der Stratten (2009), Cohen et Coughlin (2008) et Bowen et al. (2001 (2001) construisent ces indices sur la base de municipalités ou de villes. D'autres décomposent le secteur d'étude en sous-marchés associés à des quartiers (Kryvobokov, 2010, Osland and Pryce, 2012) ou encore par des limites spatiales définies sur les caractéristiques socio-économiques (Voisin et al., 2010).

Les variables de distances sont de natures continues ou binaires. Dans leur spécification continue, les distances euclidienne, de Manhattan ou réseau sont généralement les plus utilisées. Les distances binaires peuvent être interprétées comme des rayons d'accessibilité. Dans plusieurs cas, tels que ceux présentés par Samarasinghe et Sharp (2010), les stratégies de modélisation combinent les deux formes de distances, binaires et continues, pour des variables spatiales distinctes. Quelques études, telles que celles de Kryvobokov (2010), de Hui et al. (2007) ou encore de Kestens et al. (2006) utilisent la distance temps comme mesure de l'accessibilité. Parmi ces études, seuls Hui et al. utilisent des variables de distance temps binaires.

La structure de dépendance spatiale est quant à elle interprétée par la spécification d'une matrice de pondérations pouvant prendre plusieurs formes, tel que discuté dans le chapitre 3. Si certaines études intègrent des matrices de pondérations spatiales binaires (ex. : Cavailhès et al. 2009; Hui et al., 2007; Salvi, 2003), la majorité mise sur une matrice de pondérations dont la distance est considérée de manière continue. Quelle que soit la

nature de cette distance, des seuils sont postulés afin de limiter l'effet des relations spatiales à des distances allant de quelques centaines de mètres à plusieurs kilomètres. Certains de ces seuils correspondent plutôt aux limites d'un secteur ou d'un quartier (ex. : Osland, 2007). D'autres limitent les relations spatiales à un nombre de voisins prédéfini allant d'un voisin unique à une vingtaine de voisins (ex. : Nelson, 2010; Löchl et al., 2010; Salvi, 2003).

Au final, la littérature empirique présente un grand nombre d'études²⁵ de cas considérant les relations spatiales, signe d'une prise de conscience des conséquences de l'autocorrélation spatiale sur la validité des résultats. Parmi les approches proposées, plusieurs stratégies d'intégration de la dépendance spatiale sont utilisées. Pourtant, certaines de ces décisions demeurent arbitraires et continuent d'ignorer l'impact du temps. La section suivante s'attarde donc aux études qui, elles, intègrent la dimension temporelle.

4.2 ÉTUDES UTILISANT UNE APPROCHE SPATIOTEMPORELLE

Tandis que chaque propriété immobilière se définit en partie par sa localisation (Kuminoff and Jarrah, 2010), alors il est raisonnable de prétendre que chaque ménage percevra sa résidence comme étant unique (Thanos et al., 2016). Les auteurs proposent d'étendre cette logique à la localisation temporelle de sorte que les propriétés soient à la fois spatialement et temporellement situées. La section précédente a identifié un ensemble non exhaustif d'études n'intégrant qu'une seule de ces dimensions, l'espace. La présente section se concentre désormais sur des études traitant les données immobilières dans une logique spatiotemporelle (tableau 4.2).

²⁵ Pour plus de détails sur les principales conclusions de ces études voir le tableau 4.3.

Tableau 4.2: Articles considérant l'analyse spatiotemporelle

Auteurs	Marché	Secteur de l'étude	Étendue temporelle	
Can and Megbolugbe (1997)	Résidentiel	Miami (US)	1990	
Pace et al. (1998)	Résidentiel	Virginia (US)	1966	1991
Pace et al. (2000)	Résidentiel	Baton Rouge (US)	1984	1992
Tu et al. (2004)	Bureaux	Singapore (Rep. of Singapore)	1992	2001
Gelfand et al. (2003, 2004)	Résidentiel	Baton-Rouge (US)	1985	1995
Sun and Tu (2005)	Résidentiel	Singapore (Rep. of Singapore)	1990	1999
Smith and Wu (2009)	Résidentiel	Philadelphia (US)	2004	2005
Nappi-Choulet and Maury (2009)	Bureaux	Paris (France)	1991	2005
Beamonte et al. (2008, 2010a, 2010b, 2013)	Résidentiel	Zaragoza (Spain)	2002	2004
Alberto et al. (2010)	Résidentiel	Zaragoza (Spain)	2002	2004
Huang et al. (2010)	Résidentiel	Calgary (Canada)	2002	2004
Nappi-Choulet and Maury (2011)	Résidentiel	Paris (France)	1991	2005
Brandt and Maennig (2011)	Résidentiel	Hambourg (Germany)	2002	2008
Thanos et al. (2012, 2015)	Résidentiel	Athens (Greece)	1996	2001
Dubé and Legros (2013b)	Résidentiel	Lucas County (US)	1993	1998
Liu (2013)	Résidentiel	Randstad (Netherlands)	1997	2007

Plusieurs exemples de stratégies d'inclusion sont possibles pour considérer la dimension temporelle. Parmi celles-ci, la plus simple consiste certainement à inclure des effets fixes temporels. Ces effets permettent de capter certaines fluctuations temporelles du marché mais pas l'effet de dépendance structurant les observations. De plus, cette approche demeure partielle et souffre des mêmes limites que l'approche basée sur des effets fixes spatiaux mentionnée précédemment.

Le modèle STAR (Pace et al., 1998, Pace et al., 2000) est l'une des premières approches de modélisation décrites dans le chapitre précédent utilisées afin de capturer les effets de voisinage tout en respectant la chronologie de l'échantillon. Cette approche a l'avantage de diminuer radicalement le nombre de variables de contrôle tout en améliorant

le pouvoir explicatif du modèle. Le modèle STAR a donc été appliqué à plusieurs reprises dans la littérature.

Liu (2013) utilise le modèle STAR et ses conclusions démontrent que la prise en compte des dépendances spatiale et temporelle peut réduire significativement les erreurs de prédiction. Tu et al. (2004) ainsi que Sun et al. (2005) proposent une approche similaire à celle de Pace et al. (1998, 2000) mais réalisent que pour certaines observations, les processus spatiaux peuvent être d'origines multiples. Les transactions immobilières commerciales (ex. : bureaux) subissent conjointement l'influence d'une dépendance spatiale due aux effets de voisinage ainsi qu'à des effets d'immeubles. Les auteurs développent alors des modèles autorégressifs de second ordre (2STAR, 2BSTAR et 2PSTAR) comportant deux matrices spatiotemporelles destinées à capter ces deux formes de dépendances.

En revanche, Alberto et al. (2010) mettent en évidence le problème de sélection de variables associé à la logique du filtre préconisée par le modèle STAR et ses déclinaisons. En effet, même si le nombre d'effets fixes locaux est réduit, le modèle STAR implique l'inclusion d'un nombre de variables conséquent dont certaines pourraient être non significatives ou encore provoquer des distorsions dans les estimations des autres paramètres (*ibid.* : 731). Ils précisent que : « An unnecessarily high degree of model complexity, which causes a reduction in the ratio observation/number of parameters, as well as the presence of multicollinearity, might lead to a loss in estimation efficiency, misinterpretation of the results and the increasing risk of detecting specific patterns in the sample » (*idem*).

Beamonte et Gargallo (2010a), Beamonte et al. (2010a, 2008) ou encore Sun et al. (2005) répondent à ce problème par une procédure bayésienne²⁶ dont le calcul est jugé délicat par Alberto et al. (2010), qui proposent plutôt une procédure de sélection des variables. La procédure bayésienne est également utilisée par Beamonte et al. (2013) pour la construction d'indices de prix rétrospectifs pour témoigner de l'évolution des valeurs résidentielles dans une aire géographique définie.

Nappi-Choulet et Maury (2009) en font autant afin de contrôler la présence d'hétérogénéité temporelle. Alors que l'hétérogénéité spatiale a fait l'objet de nombreuses considérations, l'hétérogénéité temporelle demeure peu considérée (Nappi-Choulet and Maury, 2011). En effet, les auteurs soulignent que le modèle STAR capte les dépendances spatiales et temporelles par des coefficients autorégressifs mesurant un effet moyen sur l'ensemble de la période considérée mais que cet effet est en réalité susceptible de fluctuer. La procédure autorégressive corrige la présence d'autocorrélation par l'estimation d'un coefficient global et constant dans l'espace et le temps. Cette hypothèse de stationnarité de l'effet de dépendance est remise en question par certains auteurs. En effet, si la localisation de la résidence transigée est fixe, il est peu probable que l'effet de cette localisation sur le prix total de vente le soit également dans le temps (Alberto et al., 2010).

La régression géographiquement pondérée (GWR) ou localement pondérée (LWR) permet de répondre à ce problème de stationnarité dans l'espace. Afin d'y ajouter la dimension temporelle, Huang et al. (2010) proposent l'application d'une régression géographiquement et temporellement pondérée (GTWR). Pace et Lesage (2004) arguent que la dépendance et l'hétérogénéité spatiale doivent être modélisées simultanément et proposent l'utilisation d'un modèle autorégressif spatial local (Spatial Autoregressive Local

²⁶ L'idée générale de l'approche bayésienne est d'inférer une valeur probable à une observation en fonction d'autres estimations semblables.

Estimation, SALE). Nappi-Choulet et Maury (2011) intègrent la dimension temporelle au modèle SALE et développent un modèle autoregressif spatiotemporel local (LSTAR).

Une autre approche vise à combiner espace et temps dans une seule et unique matrice de pondérations spatiotemporelles générale (équation 3.12 et 3.13). Can et Megbolugbe (1997) seraient les premiers à spécifier une telle matrice isolant des voisins spatiotemporels (Thanos et al., 2016), limitant les relations à un rayon de 3 km et aux transactions immobilières observées dans les six mois précédents. D'autres, tels que Dubé and Legros (2013a), Smith and Wu (2009) ou encore Huang et al. (2010) appliqueront ensuite cette stratégie afin de construire une matrice de type bloc-diagonale destinée à isoler les relations spatiales par période de temps. Une fonction de distance est postulée en accord avec la logique multidirectionnelle des relations spatiales et unidirectionnelle des relations temporelles. Tout comme Can et Megboludbe (1997), Smith et Wu (2009) ainsi que Brandt et Maennig (2011) proposent d'abord la spécification d'une matrice de pondérations triangulaire inférieure.

Une fois chronologiquement ordonnées dans la matrice de pondérations temporelle carrée, les observations futures prennent place dans la partie haute de la matrice. En imposant une valeur nulle à celles-ci, les auteurs restreignent les relations spatiales irréalistes (influence du futur sur le passé) et ignorent les relations simultanées. Can et Megbolugbe (1997) préconisent l'approche basée sur la logique des ventes comparatives (comparable sales) pour construire ces indices. La démarche de ventes comparatives est basée sur le fait qu'une propriété donnée subit un effet de débordement causé par l'historique du prix des résidences voisines.

D'aucuns, tels que Huang et al. (2010) ainsi que Brandt et Maennig (2011), proposent également d'imposer un seuil temporel et spatial aux interactions. Au-delà d'une distance donnée et d'un laps de temps défini, les relations sont postulées comme étant nulles. Thanos et al. (2012 et 2015) spécifient pour leur part une matrice de pondérations

spatiotemporelle de type triangulaire inférieure comprenant une fonction de temps décroissante. La force des relations spatiales est donc en fonction de la distance physique et temporelle.

Dubé and Legros (2013a) choisissent de ne considérer que les relations entre les observations pour un mois donné. Ils comparent les résultats obtenus avec une matrice strictement spatiale et identifient clairement un effet d'une spécification strictement spatiale sur les coefficients estimés associés aux aménités des propriétés immobilières. Les auteurs constatent également une grande disparité entre les coefficients autorégressifs et, de ce fait, une évaluation erronée de l'effet marginal indirect, tel que mentionné par LeSage et Pace (2009) : « Considering the spatial dimension alone when data have an important temporal dimension can lead to the overestimation of spatial dependence in residuals and, consequently, can lead to overcorrection using the appropriate spatial econometric approaches » (Dubé et Legros, 2013a : 8).

Au regard des propositions faites pour modéliser des données en coupes transversales empilées, Thanos et al. (2016 : 87) : « are calling for a systematic approach to spatiotemporal data analysis in applied economics, formalizing this currently 'fragmented' field ». Pourtant, la majorité des travaux qui comparent les approches spatiales et spatiotemporelles ne reposent que sur un nombre limité d'échantillons ou sur des données simulées (approche Monte-Carlo). De manière complémentaire, les prochains chapitres viseront à appliquer les outils inspirés de la MA afin d'investiguer l'effet de l'omission du temps dans l'analyse d'un phénomène inscrit dans des relations de dépendance spatiotemporelles. Cette stratégie vise principalement à rejoindre deux objectifs. Elle vise d'abord à impliquer un plus grand nombre d'études empiriques dans l'analyse. Ensuite, il s'agit de recourir à des données empiriques (non simulées) inscrites dans des processus spatiotemporels réels.

Le chapitre s'est concentré sur les démarches de modélisation. Le lecteur intéressé pourra se référer au tableau 4.3 pour avoir davantage de précisions sur les conclusions principales d'études utilisant la modélisation hédonique en coupes transversales empilées dans le temps. Les prochains chapitres aborderont les études de cas et applications concrètes de la recherche.

Tableau 4.3: Revue de la littérature sur la modélisation hédonique en coupes transversales empilées dans le temps

Auteur	Résumé
Andersson et al. (2010)	Les individus sont plus dérangés par les bruits de la route que ceux du train.
Anselin et Lazano-Gracia (2008)	Les biais générés par l'omission de l'endogénéité pour des valeurs interpolées peuvent être substantiels.
Armstrong et Rodriguez (2006)	La proximité d'une voie de chemin de fer influence à la baisse les valeurs immobilières.
Bell et Bockstael (2000)	L'approche GMM autorise l'estimation avec une forme fonctionnelle de la matrice spatiale plus flexible.
Bin et al. (2008)	L'utilisation du SIG permet d'isoler avec succès l'impact des facteurs de risque des aménités spatiales.
Bowen et al. (2001)	On souligne la nécessité d'inclure le diagnostic spatial dans la spécification de modèles hédoniques.
Boxall et al. (2005)	La proximité d'infrastructures pétrolières (rayon de 4 km) en milieu rural influence les valeurs immobilières.
Brandt and Meannig (2011)	Une approche non linéaire montre que les impacts sur le prix ne sont ni constants ni proportionnels.
Brandt and Meannig (2012)	Seuls les appartements à proximité immédiate d'un groupe d'antennes en subissent les effets négatifs.
Brasington (2004)	La perte de contrôle d'une communauté sur l'offre d'un service public affecte les valeurs résidentielles.
Brasington et Samara (2008)	Lors de la vente, le type d'acte qui énonce les garanties juridiques affecte dramatiquement le prix de vente.
Can (1990)	Les modèles utilisant la méthode d'expansion sont plus précis pour refléter la réalité du marché foncier.
Capozza et al. (2005)	L'écart entre les valeurs sur le marché et de recouvrement varie en fonction des caractéristiques du bien.
Carrion-Flores et Irwin (2010)	L'utilisation de caractéristiques du sol exogènes résout le problème d'endogénéité et diminue l'ACS.
Cavallhès et al. (2009)	Le paysage importe: le paysage de nature a un impact plus significatif et les routes un impact moins significatif.
Cohen et Coughlin (2008)	Le bruit de l'aéroport influence significativement à la baisse les valeurs résidentielles.
Conway et al. (2010)	La proximité d'espaces verts possède un impact significatif sur les valeurs résidentielles.
Crespo et Grêt-Regamey (2013)	Des patrons spatiaux locaux de variabilité sont identifiés (ex.: politiques publiques ont un effet différent).
Dekkers et van der Straaten (2009)	Les prix sont affectés significativement et négativement par le bruit des avions.

Auteur	Résumé
Dermisi et McDonald (2010)	Un taux d'occupation faible des immeubles de bureaux influence les prix significativement à la baisse.
Dermisi et McDonald (2011)	L'investissement dans une certification "verte" pour des bureaux implique une plus longue période de remboursement.
Des Rosiers et al. (2010)	Il y a diminution des prix avec l'accroissement de la fréquence de passage des bus.
Donovan et al. (2007)	L'information sur les risques naturels n'a pas d'impact significatif sur le prix de vente.
Dubé et al. (2012)	L'utilisation d'une expansion polynomiale de Fourier permet de contrôler l'autocorrélation spatiale.
Fernandez-Alviles et al. (2012)	L'indice de pollution de l'air est positif et non significatif; la pollution n'a pas d'impact sur le prix des appartements.
Geoghegan et al. (2003)	La proximité d'espaces ouverts (terres) préservés accroît la valeur des propriétés.
Ham et al. (2013)	Considérer l'espace et la forêt comme étant homogènes contribue à surestimer l'impact sur les prix des propriétés.
Hansz et Hayunga (2012)	L'appartenance à un club restreint (ex.: propriétés classées historiques) augmente la valeur résidentielle.
Helbich et al. (2013)	L'hétérogénéité spatiale des prix implicites est trop complexe pour être modélisée par des indicateurs régionaux ou locaux.
Hindley et al. (2013)	La vue dégagée sur un parcours de golf influence à la hausse et significativement le prix de vente des propriétés.
Holmes et al. (2006)	Il y a impact négatif sur les valeurs résidentielles causé par un insecte exotique.
Hui et al. (2007)	La proximité du CBD et des appartements est moins corrélée, mais la WTP croît avec la proximité de la plage.
Kestens et al. (2006)	On observe un effet significatif et positif des caractéristiques du ménage sur les prix des propriétés.
Kim et al. (2003)	La pollution au dioxyde de soufre a un impact négatif sur le prix des maisons mais pas celle liée à l'oxyde d'azote.
Kim et Goldsmith (2009)	La proximité à un lieu de production porcine exerce un impact négatif.
Kostov (2009)	La régression par quantile permet l'identification d'une segmentation du marché.
Kryvobokov (2010)	Si l'ACS est contrôlée, la forme de la centralité importe peu, mais le pouvoir prédictif est en hausse avec un modèle duo centrique.
Kueth (2012)	Un petit niveau de fragmentation spatiale diminue le prix des maisons mais un haut niveau de fragmentation augmente leur prix.
Landry (2011)	Au-delà de 300 m, l'influence de la qualité d'une plage devient non significative.
Lazrak et al. (2014)	Un immeuble classé est valorisé et des impacts sont mesurés sur les maisons voisines dans un rayon de 50 m.
Leguizamon et Ross (2012)	Les résultats montrent que les individus valorisent la taille relative de la maison mais les effets absolus sont dominants.
Li et Saphores (2012)	Le développement d'espaces verts n'a pas d'impact sur les prix d'immeubles multifamiliaux.
Liao et Wang (2012)	Les prix implicites des caractéristiques des maisons varient en fonction de la distribution des prix de vente.
Löchl et al. (2010)	Les applications de simulation devraient tenir compte de la complexité de la structure spatiale du marché résidentiel.
Mihaescu et Vom Hofe (2012)	La proximité de friches industrielles exerce un impact négatif sur les valeurs résidentielles.
Militino et al. (2004)	L'étude met en évidence des différences et similitudes entre les marchés espagnols et américains.
Neelawala et al. (2013)	La WTP augmente avec l'éloignement d'une source de pollution (fonderie).
Nelson (2010)	La proximité des aménités de divertissement est fortement valorisée dans les prix de location.

Auteur	Résumé
Newburn et al. (2006)	On démontre un lien entre le prix de vente du sol et la probabilité de reconversion de l'usage du sol.
Noonan (2007)	L'analyse hédonique suggère qu'un bâtiment ou un quartier historiques sont légèrement valorisés.
Osland et al. (2007)	Différentes fonctions du temps de trajet sont testées; la bonne spécification de la fonction diminue l'ACS.
Osland et Pryce (2012)	On procède à l'identification de preuves convaincantes à l'effet que la mesure de l'accessibilité est non monotone.
Osland et Thorsen (2013)	Les résultats encouragent à tenir compte des sous-centres situés à proximité du CBD.
Osland. (2010)	L'étude présente un exemple pratique de modélisation hédonique.
Parent et Vom Hofe (2013)	Les processus spatiaux améliorent les résultats; le SEM capture mieux les effets de débordements que le SAR.
Paterson et Boyle (2002)	La vue influence le prix de vente et son omission peut invalider les résultats des variables environnementales.
Patton et McErlean (2003)	Le prix du sol agricole est déterminé par ses caractéristiques mais aussi par celles des terres voisines.
Perdomo (2011)	La ligne de bus rapide a un impact positif sur le prix de vente.
Price et al. (2010)	Le prix des maisons baisse lorsque des arbres à proximité sont affectés par des insectes nuisibles.
Salvi (2007)	L'article rapporte un effet modéré de l'impact du bruit sur les valeurs résidentielles.
Samarasinghe et Sharp (2010)	Une maison située en zone inondable est vendue moins cher qu'une maison semblable en dehors d'une telle zone.
Sander et Haight (2012)	Plusieurs aspects environnementaux esthétiques ont un impact significatif sur le prix de vente.
Saphores et Li (2012)	Les résidents de Los Angeles veulent plus d'arbres mais ne sont pas disposés à payer pour eux.
Sedgley et al. (2008)	Il y a un impact significatif des résultats scolaires sur le prix de vente des propriétés.
Seya et Tsutsumi (2013)	L'étude suggère une spécification en deux temps de la sélection du modèle: 1) matrice W; 2) variables exploratoires.
Tsutsumi et Seya (2009)	Les estimations (<i>estimated benefits</i>) sont inégales entre processus spatiaux et erreurs autorégressives et pour isotropie et anisotropie.
Wallner (2012)	La vue sur l'eau ou sur des terres a un impact positif sur le prix de vente des résidences.
Walsh et al. (2011)	L'étude identifie des impacts positifs et significatifs des effets de voisinage dans l'industrie du vin.
Wilhelmsson (2002)	Un usage non critique des méthodes économétriques spatiales peut poser problème pour interpréter les paramètres.
Yu et al. (2012)	Les ménages valorisent la proximité à un centre d'achats dans un rayon de trois à vingt minutes de transport.
Zhu et al. (2011)	La dépendance anisotropique peut réduire les erreurs de prédiction dans et en dehors de l'échantillon.

CHAPITRE 5

L'INFLUENCE DU TEMPS SUR LA DÉPENDANCE SPATIALE : UNE MÉTA-ANALYSE BASÉE SUR L'UTILISATION DE LA MODÉLISATION HÉDONIQUE EN IMMOBILIER

5.1 PRÉSENTATION DE L'ARTICLE

Le premier article de la thèse, intitulé « About the Influence of Time on Spatial Dependence: A meta-analysis using real estate hedonic pricing models », a été rédigé en collaboration avec le professeur Jean Dubé. L'article a d'abord fait l'objet d'une présentation dans le cadre du congrès annuel de l'*American Real Estate Association* à San Diego, CA (É.U.) en 2014. Suite à cette présentation, le travail a été soumis à la revue *Journal of Real Estate Literature* sur la proposition de Madame Sophia Dermisi et a été accepté pour publication dans sa version définitive en 2016. La recherche a été subventionnée par le Fonds de recherche sur la société et la culture (FQRSC).

Les données immobilières se particularisent par leur caractéristique désagrégée et non répétée à intervalle fixe (Huang et al., 2006). Dans ce contexte, il n'est pas rare que ces données, pourtant inscrites dans des schémas de dépendance spatiotemporelle, soient considérées dans une perspective exclusivement spatiale. Ainsi, l'article vise à évaluer l'influence de l'omission de la dimension temporelle dans la modélisation hédonique immobilière sur certaines statistiques spatiales.

Certaines études précédentes ont abordé la problématique (Dubé and Legros, 2013a, Dubé and Legros, 2014a, Smith and Wu, 2009). Néanmoins, celles-ci utilisent généralement un nombre restreint d'échantillons ou des données simulées. L'idée générale de l'article est de recourir à la méta-analyse afin de permettre une synthèse de travaux empiriques. Les statistiques spatiales considérées sont l'indice I de Moran (détection de la dépendance spatiale) et les coefficients autorégressifs liés aux modèles économétriques spatiaux, λ (SEM) et ρ (SAR) (correction de la dépendance spatiale). Les résultats suggèrent que la mesure et la correction de la dépendance spatiale sont effectivement affectées par la dimension temporelle de diverses manières.

5.2 ABOUT THE INFLUENCE OF TIME ON SPATIAL DEPENDENCE: A META-ANALYSIS USING REAL ESTATE HEDONIC PRICING MODELS

5.2.1 Abstract

Real estate transactions data usually consist of a collection, over time, of spatial layers of individual transactions. Despite the presence of a temporal dimension, it is not uncommon for these data to be analyzed using a strictly spatial approach. In this paper, we assess the impact of the temporal omission on spatial statistics (Moran's I and autoregressive parameters) when unrepeatably collected data are used. Using a meta-analysis approach, a sample of empirical studies is collected. The results show that spatial dependence measurement and correction is not an obvious task and that time affects these statistics in different ways.

5.2.2 Introduction

The increasing availability of geolocated data opens up new avenues to spatial analysis, especially in real estate applications. Following the identification of the spatial dependence *issue* by Cliff and Ord (1968), the number of scientific publications associated with the keywords *spatial autocorrelation*, *spatial analysis* and *spatial econometrics* has increased since the 90's and illustrates the enthusiasm for such methods (Figure 5.1).²⁷

²⁷ Results are from the Scopus search engine in December 2015. Scopus is a scientific database comprising 19 500 online journals from 5000 different publishers for a total of 47 million records. Of these references, over 26 million are from the 1996 to 2014 period and 21 million from the 1826 to 1996 period (www.info.sciverse.com/scopus/scopus-in-detail/facts).

While the classical correlation statistic refers to the relations between variables, spatial autocorrelation refers to the correlation between values of different observations for a single variable and is described as the coincidence between observed variables depending on location (LeSage and Pace, 2009; Chasco and Lopez, 2008; Anselin and Bera, 1998). Spatial autocorrelation is a key concept in studies using spatial data: "No other concept in empirical spatial research is as central to model building as is spatial autocorrelation" (Getis, 2008: 299).

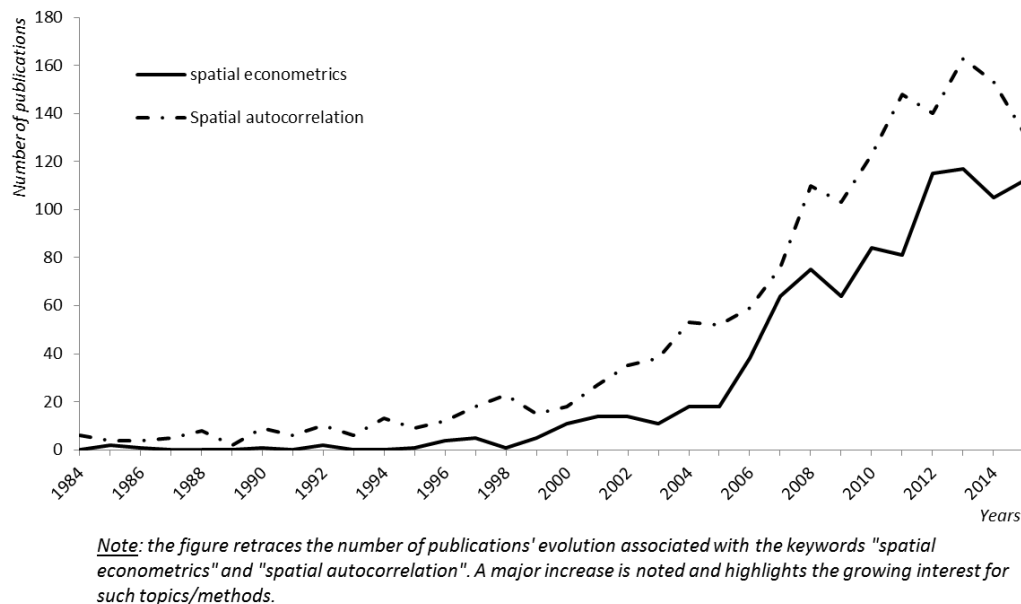


Figure 5.1: Progression of research papers' publications (1984-2015)

Spatial analysis and spatial econometrics have been developed after the pioneer work of Cliff and Ord (1968) and Paelinck and Klassen (1979) and are a response on how to account for spatial dependence in statistical models. The difference between the two lies in

how spatial dimension is accounted for in modelling strategies, especially through the *data generating process* (DGP) and the way we specify the spatial connections among geographical observations. In spatial econometrics, this connection is formalized through a weights matrix. Some argue that even in an exclusively spatial perspective, the weights matrix is the cornerstone of spatial econometrics, and due to its exogenous specification, also its main weakness (Bhattacharjee et al., 2012; Chen, 2012; Elhorst, 2010; Anselin, 2002). On the other hand, others claim that too much attention is drawn to defining *the* good spatial weight matrix specification, which became the biggest myth in spatial econometrics (LeSage and Pace, 2010).

In recent years, spatial econometric developments tend to include the temporal dimension to the spatial cross sectional perspective (Arbia, 2011). Adding the temporal perspective is a critical transition to analyze the dynamic process and its evolution (Anselin, 2002). By including a temporal dependency in spatial analysis the DGP becomes more complex. Spatiotemporal models are mainly developed through the panel or pseudo-panel perspectives (Elhorst, 2013). Few attempts have been made, outside the spatial panel analysis perspective, to combine spatial and temporal dependencies for spatial data collected over time, while not being repeated, such as in the case of real estate transactions. In such a case, time constraints are often neglected, assuming the simultaneous observation of the data (Dubé et al., 2011a, Dubé and Legros, 2013a). However, this assumption remains to be verified (Dubé and Legros, 2013b) and a misinterpretation of the data structure could affect the detection and adequate correction of spatial autocorrelation, since ignoring the temporal dimension suggests an overestimation of spatial relations.

To our knowledge, no study has attempted to empirically demonstrate the impact of such omission on the spatial dependence statistics beyond a limited number of case studies or Monte Carlo simulations (see for example Dubé and Legros, 2015, Dubé and Legros, 2014b). The stakes are high because a poor detection and correction for autocorrelation

may bias estimators or affect their variance and therefore invalidate the results obtained from statistical analysis. The presence of spatial autocorrelation in residuals of ordinary least square (OLS) models leads to increased risk of type 1 error and affects the model goodness-of-fit statistic (R^2) (Haining, 2009; Anselin, 2002): "This is problematic since failing to account for serial and spatial autocorrelation when present causes the OLS estimators to lose their property of efficiency" (Elhorst, 2001: 119). In the context of real estate analysis, erroneous conclusions mean wrongly measuring the impact of some environmental amenities on house prices and, in consequence, badly measure the economic value of urban externalities on real estate buildings. In some particular cases, it could even suggest the implementation of unproductive development or landscaping policies.

The paper aims to assess the impact of the omission of the temporal dimension on spatial dependence statistics and parameters. First, the detection of spatial autocorrelation (the Moran's Index) can be affected. Second, the estimated autoregressive coefficients used to control for an endogenous (spatial autoregressive model - SAR, ρ) or exogenous (spatial error model - SEM, λ) spatial model, or data generating process, can be biased. Using a meta-analysis approach, these statistics/parameters are collected over a sample of empirical studies related to real estate applications. The temporal dimension, the data generating process (DGP), the data characteristics, the weights matrix specification or the model specifications are some of the considered factors likely to affect spatial autocorrelation measures and correction. Three models are developed to link the amplitude of the spatial statistics to the characterization of the studies.

The results show a limited impact of the weights matrix specification and evidence of the influence of the temporal dimension on spatial statistics. However, the results show that such effect is different according to the spatial statistics and the research highlights some distinct behaviour on the autoregressive parameters. Overall, spatial autocorrelation detection and correction is mostly affected by sample characteristics (number of

observations, spatial and temporal density). The two autoregressive terms (lagged error term and lagged dependent variable) suggest different temporal processes fluctuations over space. In light of previous results, this may suggest a better understanding of the data in specialized literature or different estimation habits and paper review process. The impact of the results are of prime importance because no universal DGP and weights matrix specification rules exist, thus a solid knowledge and understanding of the data generating process is required in empirical analysis.

The paper is divided into five sections. The first section discusses the data structuring processes and the consequences of a strictly spatial analysis for spatially and temporally structured data. The second section presents the main econometric strategies used for spatial dependence detection and correction. Moran's *I* and autoregressive specifications are exposed. The third section is devoted to the research methodology. The meta-analysis is defined as well as the model, the sample of studies used and the main research hypothesis. Following this, the results are presented then a discussion and a conclusion follow.

5.2.3 Spatial and temporal dimensions in real estate data

Spatial dependence among observations is now widely recognized in the literature (Anselin, 2010; Cliff and Ord, 1969) and has been considered in real estate studies since the 90's (Can, 1992, Dubin, 1998). The hedonic pricing model (HPM) allows explaining real estate price variations as a function of intrinsic and environmental characteristics. Location is a major determinant of real estate values. Being spatially situated, those values incorporate spatial dynamics influences through neighbourhood, accessibility or environmental characteristics (Bramley et al., 2008, Des Rosiers, 2002, Hite et al., 2001, Mohammad et al., 2013). In such a context, it is now widely recognized that standard

ordinary least square (OLS) regression models suffer from residual spatial autocorrelation violating the error terms independent hypothesis (Dubé and Legros, 2014a).

To deal with these spatial processes, spatial econometrics proposes to develop an exogenously specified spatial weights matrix to formalize the spatial relations among observations. Two types of spatial weights matrices are mainly used: i) a binary specification indicating whether the observations are spatial neighbours (or not); and ii) a general distance function reflecting the decreasing effect with respect to increasing distance, such as inverse distance (or squared distance) function or negative exponential function.²⁸

By integrating a weights matrix, models with spatial interactions manage the relations between observations that may lead to collective behaviour or aggregation patterns in space and that emphasize the role of the location (Anselin, 2002). The cross sectional analysis does not identify the precise nature of the causal mechanisms of such processes (Gibbons and Overman, 2012). Including the temporal dimension in the analysis of spatial phenomena allows the observations to be chronologically organized and contributes to the identification of dynamic processes (Dubé and Legros, 2014c). The econometrics' attention has long been focused on time series, and the possible relations of dependence between the cross sections have long been ignored in the analysis panel (Anselin and Griffith, 1988). The repetition of observations over time allows defining the dependency and spatiotemporal causality processes structuring data through spatial panel data (Arbia, 2011; Elhorst, 2014).

Yet, some data may not be collected continuously, making it impossible to use a conventional (spatial) panel analysis. This is usually the case of microdata, such as real estate values, characterized by a disaggregated level. In the case of real estate data, Huang

²⁸ See Chen (2012) for a complete discussion on the impact of choosing a particular specification.

et al. (2006) note that the low number of repeated observations compromises a time-series analysis. The authors then propose to treat them as a pseudo-panel, which differs from the *classical* panel analysis by aggregating groups of observations not repeated on similar and invariable characteristics over time and/or space (Gardes et al., 2005). The pseudo-panel approach allows postulating the temporal and spatial process but causes losses of information due to aggregation, and confronts the scholar with the modifiable areal unit problem (MAUP, see Openshaw, 1983). Aggregation acts as a filter on which the collected values depend and may generate a scale problem: at different levels of spatial aggregation exist different factors responsible for observed results (Haining, 2009). Anselin (2002) also reiterates the dangers of inference based on aggregate data (risk of ecological fallacy). Moreover, the choice of the unit of observation is central in spatial processes analysis and can foster the presence of spatial autocorrelation (Anselin and Griffith 1988; Anselin, 1988). Also, "at an aggregate level, heteroskedasticity should be expected and an *i.i.d.* assumption for the errors is incompatible with the aggregation rule" (Anselin, 2002: 261).

Hedonic analyses usually avoid these spatial aggregation problems but ignore the temporal dimension and often analyze data recorded in space *and* time using an exclusively spatial perspective. This approach amounts to an aggregation of data, not in space but in time, forcing spatial relations among observations temporally distant, regardless of the period of their measurement.

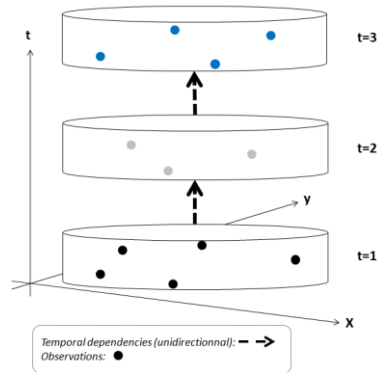
However, the simultaneous consideration of time and space presents some challenges. One of these is the different nature of spatial and temporal relationships. As shown in Figure 5.2, spatial relationships are multidirectional, so that each observation maintains reciprocal links with its neighbours (Figure 5.2a). Temporal relations are unidirectional so that future transactions are influenced by past transactions, while the reverse is impossible (Figure 5.2b). Ignoring the latter logic, scholars falsely assume a share of non-existent relations and therefore overstate the spatial ones. By omitting the temporal dimension, the

assumption of simultaneity of observation is implicitly made. Therefore, bias can be introduced in the detection and correction of spatial autocorrelation by econometric procedures leading to erroneous conclusions.

Different works outlined the risk of introducing bias in autoregressive coefficients when the analysis lacks the temporal dimension (LeSage and Pace, 2009; Dubé and Legros, 2013a; 2013b). Some propositions were made in order to consider temporal and spatial constraints for spatial data pooled over time. Pace et al. (1998; 2000) suggested a strategy to combine spatiotemporal dependencies by the matrix product of a spatial and temporal independent weights matrix. The proposed STAR model includes *filters* to model neighbourhood effects by respecting the sample chronology (similar properties recently traded). This approach helps to drastically reduce the number of control variables while improving the explanatory power of the model. Liu (2013) also advocates a spatiotemporal autoregressive approach in predictive analysis; the findings show that taking into account the spatial and temporal dependencies can significantly reduce prediction errors.

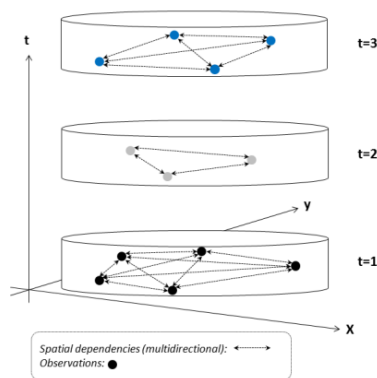
This framework was subsequently used for developments in spatiotemporal analysis (Nappi-Choulet and Maury, 2009 and 2011; Beamonte et al. 2008; 2010a; 2010b) or empirical applications (Tu et al., 2004; Sun et al., 2005). However, this approach fails to directly capture spatiotemporal effects (Smith and Wu, 2009; Dubé and Legros, 2014a). Moreover, Nappi-Choulet and Maury (2009; 2011) note that because the autoregressive procedure corrects the presence of autocorrelation as a global and constant coefficient in space and time, it captures the spatial and temporal dependencies as an average effect for the whole considered period, even if this effect is likely to fluctuate. The stationary assumption of the dependency effect is questioned by some authors, who argue that if the location of the traded property is fixed, it is unlikely that the effect of this location on the total selling price is also stable over time (Alberto et al., 2010).

Figure 5.2a



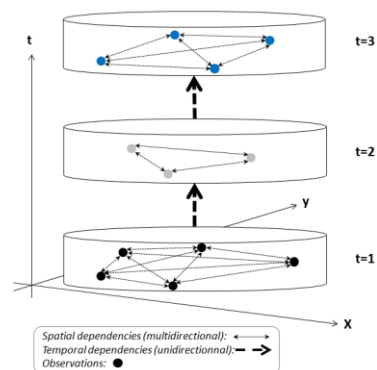
- ❖ The first illustration (2a) presents the unidirectional temporal relations between strata of observations (sales) temporally delimited. Each temporal layer contains different (unrepeated) observations. In this design, the spatial dependence is ignored.

Figure 5.2b



- ❖ The second illustration (2b) presents the same strata of unrepeated observations (sales) spatially interrelated to capture spatial dependence. In this design the temporal dependence between layers of observations is ignored. Consequently, the analyst assumes that all sales are simultaneous.

Figure 5.2c



- ❖ The third illustration (2c) suggests a way to simultaneously consider spatial and temporal dependencies for unrepeated data. Such an approach conserves the unidirectional and multidirectional nature of temporal and spatial dependencies.

Figure 5.2: Spatial and temporal dependency relations

Others proposed the specification of a general spatiotemporal weights matrix (Dubé and Legros, 2013a; Smith and Wu, 2009; Huang et al., 2010). The latter proposal has the advantage of freeing scholars from the aggregation and modifiable areal unit problems (MAUP). But combining temporal and spatial processes is not an obvious task (Figure 5.2c). Smith and Wu (2009) suggest building a lower triangular spatial weights matrix. With data chronologically organized, this approach allows to ignore the unrealistic temporal relations, i.e., the future influencing the past, but not to account for simultaneous observations. Huang et al. (2010) suggest creating such spatiotemporal weights matrices truncated in space and time. Dubé and Legros (2013a) propose to account for simultaneous transactions only. Dubé and Legros clearly identified that a spatial approach has impacts on the coefficients associated with housing amenities and creates a wide disparity between autoregressive parameters and thus an erroneous evaluation of the indirect marginal effect as suggested by LeSage and Pace (2009).

5.2.4 Dealing with spatial autocorrelation: detection and correction

Spatial autocorrelation can emerge from a model misspecification due to omitted variables, such as independent variables hard to measure and measurement errors, or the use of incorrect functional form (Le Gallo, 2002; Osland, 2013; McMillen, 2010). As underlined by McMillen (2010: 138), "the main difficulty with modelling spatial data is the combination of non-linearity and missing variables correlated over space." This omission of significant variables introduces bias on the estimated coefficients, while spatial heteroskedasticity, through correlation of the error terms, invalidates the use of the usual statistical test (Anselin and Arribas-Bel, 2011).

The spatial weights matrix specification is still a challenge in empirical investigation (Fingleton, 2003, Getis, 2009): it may generate bias in autoregressive coefficients when

misspecified (LeSage and Pace, 2009) or over-connected (Smith, 2009). The production of statistical detection of spatial dependence and estimated autoregressive statistics relies on the weights matrix specification. So far, it has been shown that the use of a spatial matrix in a spatiotemporal context influences spatial statistics (Dubé and Legros, 2014b; 2013a; 2013b).

Using existing empirical studies, the research proposes to determine the consequences of the omission of time on detection (Moran's I) and control (autoregressive coefficients) of spatial dependence. The Moran's I index (Moran, 1948; 1950) is the most used statistic to detect such patterns (Le Gallo, 2002; Griffith, 2012). Applied to the error terms obtained from an ordinary least squares (OLS) model, this index is an auto-covariance function between the error term, ε , and its spatial lagged variable, $\mathbf{W}\varepsilon$. The spatial weights matrix, \mathbf{W} , is row-standardized, with the sums of the elements on a row equal to one. The Moran's I index is used to determine whether data are randomly distributed over space (Equation 5.1 - Haining, 2009).

$$I = \frac{\varepsilon' \mathbf{W} \varepsilon}{\varepsilon' \varepsilon} \quad (5.1)$$

When such statistics appear to be significant (different from zero), the statistical models need to account for spatial dependence. In real estate applications, there are two main ways of incorporating spatial dependence in an econometric model. The first way is to assume that spatial dependence occurs because a spatial spillover effect influences the DGP. The spatial autoregressive (SAR) model assumes that there is an endogenous spatial process relating to an influence of neighbouring house transaction prices on a given sold property. A spatial lag of the dependent variable is included in the deterministic part of the model to integrate spillover effects (Equation 5.2).

$$y_{it} = D_{it}\delta + \mathbf{W}y_{it}\rho + X_{it}\beta + \varepsilon_{it} \quad (5.2)$$

Where the matrix X_{it} refers to the intrinsic and environmental characteristics of the real estate good and β the associate coefficients expressing the (mean) implicit (hedonic) prices of amenities (see Small and Steimetz, 2012, Steimetz, 2010); $\mathbf{W}y_{it}$ is the lagged dependent variable for each observation i at the period t and the spillover effect is measured by the coefficient ρ . The D_{it} matrix represents a series of temporal dummy variables for the property transaction period and δ refers to the associated coefficients capturing sample difference for each time period (Wooldridge, 2010). This model allows modelling the dependent variable for a spatial unit based on the same variable for neighbouring spatial units (Elhorst 2010). In other words, this specification allows distinguishing the effect of the independent variable in two direct and indirect marginal effects (LeSage and Pace, 2009).

The second way of dealing with spatial dependence among residuals is to assume that the error terms are spatially autoregressive. The spatial error model (SEM) refers to a lag of the stochastic variable suggesting the omission of spatially correlated variables specific to neighbourhood effects (Equation 5.3) (Bivand, 1998).

$$\begin{aligned} y_{it} &= D_{it}\delta + X_{it}\beta + u_{it} & 5 \\ u_{it} &= \lambda\mathbf{W}u_{it} + \varepsilon_{it} & (5.3) \end{aligned}$$

The exogenous spatial dependence process is captured through spatial/environmental unobserved characteristics that encourage similar behaviour (Elhorst 2010). Unlike the SAR model, the coefficients obtained using an OLS approach are not biased.²⁹ However,

²⁹ Unless the latent spatial structure is correlated with the independent variables.

the estimated variance obtained using the OLS approach is not efficient and therefore affects the usual statistical tests, such as the significance tests (LeSage and Pace, 2009; Dubé and Legros, 2013a).

In all cases, the structure of the weights matrix, \mathbf{W} , can clearly have an influence on the amplitude of the detection statistics, through the Moran's I index, as well as on the amplitude of the autoregressive coefficients, mainly for spatiotemporal data. The existing analysis suggests that considering the spatial dimension alone creates divergence in some explanatory coefficients and generates a wide disparity between autoregressive parameters, and thereby an erroneous assessment of the indirect marginal effect (LeSage and Pace, 2009). Various approaches are proposed to capture the spatial autocorrelation for unrepeated spatial data collected over time. However, given the complexity of the proposed approaches, many studies still ignore the temporal dimension and impacts of such omission. Based on a single case study, Dubé and Legros (2013a) found that a spatiotemporal matrix tends to produce a higher variance compared to a strictly spatial matrix. Considering the spatial dimension alone, there is a risk of overestimating the spatial dependence and thereby to overcorrect it thereafter with the autoregressive specification.

5.2.5 Methodology: The meta-analysis and methodological framework

A meta-analysis (MA) is proposed to assess the consequences of the omission of the temporal dependence on statistical correction and detection of spatial autocorrelation. The MA is an appropriate method in order to conduct a comprehensive summary of the results in a context of abundant scientific knowledge (Florax et al., 2002). This approach is defined as a statistical analysis to the results of studies that, in the case of the economy, estimate a particular strength (Schipper et al., 2012). The term MA is introduced by Glass (1976) as a

rigorous statistical analysis of a collection of individual studies' results in a rapidly expanding research literature (Glass, 1976 in Florax et al., 2002).

Originally, the MA is based on a desire for knowledge synthesis and is largely applied in health, psychology or education fields of research, while the MA seems more timidly applied in the context of real estate analysis. Yet, in a context of incremental theory building and based on knowledge accumulation, the MA marks a pause and a time for reflection in the race for the production and accumulation of scientific results. There are still few study examples incorporating the MA in real estate analysis to synthesize study results in particular to consider the effect of transportation infrastructures on property values where case studies are abundant. Debrezion et al. (2007) and Mohammad et al. (2013) synthesized research results considering the impact of railway stations and infrastructures on land, residential and commercial property values. Airports and aircraft noise were also considered by Nelson (2004) and Schipper et al. (2002).

Scientific studies are produced in a variety of contexts and with different research strategy choices, which makes them all the more difficult to synthesize. By considering data collected over space and time in a strictly spatial perspective or by exogenously specifying a weight matrix specification, the scholar makes some strategic data generating process (DGP) choices. It seems legitimate to question the consequences of those choices, and the MA appears to be an adequate method to answer this question. For a sample of selected studies, the proposed strategy is to explain changes in spatial autocorrelation detection statistics (Moran's I) and correction statistics (autoregressive parameters ρ and λ) using different explanatory factors (Equation 5.4).

$$Y_{mi} = \sum_{k=1}^K DGP_{mi} \theta_{mi} + (\gamma_{mi} T_{mi} + \phi_{mi} \ln T_{mi}) + \sum_{l=1}^L Study_{mi} \mu_{mi} + v_{mi} \quad (5.4)$$

Where Y_{mi} refers to the explained spatial statistic (Moran's I) or parameters (λ and ρ) of study i . The matrix DGP_{mi} is a set of variables controlling for the model characteristics (specification and weights matrix), and the matrix $Study_{mi}$ is a set of variables describing the research and dataset characteristics and v_{mi} is an independently and identically distributed (*iid*) error term. The details of the included variables are presented in Table 5.2.

A flexible non-linear specification of the effect of the temporal dimension on the spatial statistics amplitude is then proposed. Introducing T_{mi} and $\ln T_{mi}$ in the equation, representing, respectively, the temporal and the *log* of the temporal span of the dataset (considered in month), one may consider a full set of possibilities. According to this specification, the form of the temporal dataset length effect can be strictly decreasing if the parameters $\gamma < 0$ and $\phi < 0$ are both negative. Inversely, the effect can strictly be positive if $\gamma > 0$ and $\phi > 0$. For diverging signs, the obtained curve is going to be convex ($\phi > 0$ and $\gamma < 0$) or concave ($\phi < 0$ and $\gamma > 0$). The optimal value is obtained by the ratio $(-\phi)/\gamma$.

Based on Equation 5.4, the net effect of the dataset's temporal range is calculated. The effect the time (in number of months) has on spatial statistics is given by the partial derivative of the spatial autocorrelation detection equation. The effect of the temporal range on the spatial statistic of interest can be determined by:

$$\frac{\partial Y_{mi}}{\partial T_{mi}} = \gamma_{mi} - \frac{1}{T_{mi}} \phi_{mi} \quad (5.5)$$

Thus, taking the significant parameters, it is possible to evaluate the effect of time on the amplitude of the spatial statistics. Moreover, the influence of time can also be measured through other variables that are related to the way temporal dimension is accounted for in HPM, notably using temporal dummy variables.

As always, the estimated coefficients, ϕ , θ , μ , γ , measure the effect of the independent variable on the dependent variable. The study aims to test three research assumptions and is structured according to a hypothetical-deductive approach to assess the impact of the temporal dimension on targeted spatial statistics/parameters (Moran's I and autoregressive coefficients ρ and λ).

1. Omission of the time significantly influences spatial statistics in empirical analysis. However, the magnitude of such impact varies depending on the model's *data generating process* (DGP).
2. The temporal span of the dataset, i.e., the elapsed time between the first and the last observation, affects the spatial statistics. The wider the temporal span is, the more connected the observations are. In fact, it is assumed that with the exclusion of time, all the observations are related in space independently from their chronology. For this reason, observations may be over-connected which is likely to affect spatial statistics, as exposed by Smith (2009).
3. Data and study characteristics such as the number of observations, the study location or the study purpose affect the spatial autocorrelation statistics.

5.2.6 Data collection

The analysis is based on empirical study results. A collection of 85 studies was identified among a large set of scientific papers. The literature review was conducted in three stages. A wide range of studies considering real estate values (unrepeated sales only) through the prism of the hedonic modelling was first identified.

Initially, extensive research was conducted over different database search engines such as *Scopus*, *Elsevier*, *Ebscohost*, *ScienceDirect*, *SpringerLink*, *Wiley*, *JSTOR*, or even *Scholar Google*. Over 300 papers corresponding to keywords “hedonic pricing model”; “spatial autocorrelation”; “hedonic analysis” and associated with real estate analysis were identified.

The second step aimed at sorting the collected papers. Of these selected studies, only the ones that report Moran’s *I* statistics and/or estimated spatial autoregressive coefficients were identified and retained. At the same time, only papers with precise and accurately reported methodology on the construction of the weights matrix and on the DGP were retained.

Finally, particular attention has been paid to the bibliography of papers preserved to identify some possible missing studies. New papers were identified and added using a *snowball* selection for which the whole process has been repeated until information redundancy.

The collected studies consider various issues. Some of them attempt to outline the effect of certain externalities on residential values, while others use real estate data for methodological purposes. The meta-analysis uses papers collected in 47 different journals (see Table 5.1).³⁰

The meta-regression models consider three particular types of variables referring to the dataset characteristics, the data generating process and the model estimation method (Table 5.2). The dataset characteristics are used to describe the context in which data are

³⁰ Unpublished papers were rejected from the data collection procedure. However a few unpublished observations were kept and included in the analysis because of their repeated presence in the selected papers’ bibliography. These redundant occurrences outlined quality of the analysis and methodology description. These exceptions are considered relevant in this analysis.

collected as well as the sample features. The context refers to the location of the collected observations, distinguishing North American samples from other places (mainly Europe and Asia). Considering these urbanization specificities, city density or political factors may be particular to each location, we assume that such factors may affect residential values or spatial dependence processes. Considering that real estate data may be used for traditional case studies or methodological purposes, both approaches are distinguished. The dataset's number of observations is included in our meta-regression models specification as well as the dataset time range.

Table 5.1: Considered journals

Scientific journals	
1 - <i>Agricultural and Resource Economics Review</i>	25 - <i>Journal of Geographical Systems</i>
2 - <i>American Journal of Agricultural Economy</i>	26 - <i>Journal of Housing Economics*</i>
3 - <i>Applied economic Letters</i>	27 - <i>Journal of Property Investment and Finance *</i>
4 - <i>Australian Journal of Agricultural and Res. Eco.</i>	28 - <i>Journal of Property Research*</i>
5 - <i>Building and Environment</i>	29 - <i>Journal of Real Estate Finance and Economics*</i>
6 - <i>Ecological Economics</i>	30 - <i>Journal of Real Estate Literature *</i>
7 - <i>Economic Development Quaterly</i>	31 - <i>Journal of Real Estate Portfolio Management</i>
8 - <i>Economic Geography</i>	32 - <i>Journal of Real Estate Research*</i>
9 - <i>Empirical Economics</i>	33 - <i>Journal of Regional Analysis and Policy</i>
10 - <i>Environment and Planning</i>	34 - <i>Journal of Regional Science</i>
11 - <i>Environmental and Resource Economics</i>	35 - <i>Journal of Statistical Planning and Inference</i>
12 - <i>Federal Reserve Bank of St-Louis-Work papers</i>	36 - <i>Journal of Transport and Land Use</i>
13 - <i>Forest Policy and Economics</i>	37 - <i>Land Economics</i>
14 - <i>Growth and Change</i>	38 - <i>Landscape and Urban Planning</i>
15 - <i>Housing Studies*</i>	39 - <i>Real Estate Economics*</i>
16 - <i>Housing Theory and Society</i>	40 - <i>Regional Science and Urban Economics</i>
17 - <i>International Journal of Forecasting</i>	41 - <i>Resource and Energy Economics</i>
18 - <i>International Journal of Housing Market</i>	42 - <i>Review of Economics and Statistics</i>
19 - <i>International Journal of Sustainable Transp.</i>	43 - <i>Spatial Economic Analysis</i>
20 - <i>Journal of Agricultural Economics</i>	44 - <i>The Annals of Regional Science</i>
21 - <i>Journal of Environmental Eco. and Management</i>	45 - <i>Transportation</i>
22 - <i>Journal of Environmental Management</i>	46 - <i>Transportation part D</i>
23 - <i>Journal of Forecasting</i>	47 - <i>Unpublished paper</i>
24 - <i>Journal of Geographic Systems</i>	48 - <i>Urban Studies</i>

* Academic journals specialized in real estate analysis.

Table 5.2: List of the considered variables in the meta-regression models

Variable	Definition	Type
Number of observations	<i>Number of observations in the sample from which the statistic is computed (log)</i>	<i>Continuous</i>
Study location: North America	<i>The data is collected in North America</i>	<i>Binary</i>
Study location: Other	<i>The data are collected outside North America. Ex: Europe, Asia, etc. (reference)</i>	<i>Binary</i>
Study focus : methodological purpose	<i>Real estate data are used for methodological considerations</i>	<i>Binary</i>
Study focus: Case study	<i>Real estate data are used in order to consider a particular case study (reference)</i>	<i>Binary</i>
Presence of continuous distance variables	<i>Some of the spatial variables are considered as continuous variables</i>	<i>Binary</i>
Presence of buffer distance variables	<i>Some of the spatial variables are considered as buffer variables</i>	<i>Binary</i>
Independent variable: proximity variables	<i>Number of spatial variables included in the model</i>	<i>Continuous</i>
Independent variable: other	<i>Number of non-spatial variables included in the model</i>	<i>Continuous</i>
Functional form: Linear	<i>The specified model has a linear functional form</i>	<i>Binary</i>
Functional form: semi-log	<i>The specified model has a semi-log functional form</i>	<i>Binary</i>
Functional form: Log-log	<i>The specified model has a double-log functional form (reference)</i>	<i>Binary</i>
Weight matrix: binary	<i>The weights matrix used is binary</i>	<i>Binary</i>
Weight matrix: number of neighbours	<i>The weights matrix is based on a nearest neighbour specification</i>	<i>Binary</i>
Weight matrix: with a threshold	<i>The weights matrix is specified considering a distance cut-off</i>	<i>Binary</i>
Weight matrix: inverse distance	<i>The continuous weights matrix is specified using an inverse distance function</i>	<i>Binary</i>
Weight matrix: squared inverse distance	<i>The continuous weights matrix is specified using a squared inverse distance function</i>	<i>Binary</i>
Weight matrix: Other distance function	<i>The continuous weights matrix is specified using another distance function (reference)</i>	<i>Binary</i>
Weight matrix: spatio-temporal	<i>The weights matrix is specified considering spatial and temporal dimensions</i>	<i>Binary</i>
Dataset temporal range	<i>Number of months elapsed between the first and the last observations</i>	<i>Continuous</i>
Dataset temporal range	<i>Log of the number of months elapsed between the first and the last observations</i>	<i>Continuous</i>
Presence of temporal fixed effects	<i>The model includes temporal fixed effects controlling for the period of the transaction</i>	<i>Binary</i>
Dataset temporal range*Real estate	<i>Real estate specialized journals: Number of month elapsed between the first and the last observations</i>	<i>Continuous</i>

Dataset temporal range*Real estate	<i>Real estate specialized journals: Log of the number of month elapsed between the first and the last observations</i>	<i>Continuous</i>
Presence of temporal fixed effects*Real estate	<i>Real estate specialized journals: The model includes temporal fixed effects controlling for the period of the transaction</i>	<i>Binary</i>
Estimation: GM	<i>The autoregressive model is estimated using the generalized method of moments</i>	<i>Binary</i>
Estimation: TSLS	<i>The autoregressive model is estimated using the two-stage least squares method</i>	<i>Binary</i>
Estimation: ML	<i>The autoregressive model is estimated using the maximum likelihood method (reference)</i>	<i>Binary</i>
Model: SAC	<i>The autoregressive model integrates both endogenous and exogenous spatial processes</i>	<i>Binary</i>

The data generating process (DGP) refers to the modelling, and theory-based choices made by the scholars. These choices involve the functional form of the model or the number and type (spatial or non-spatial) of independent variables. The models control for the presence of binary variables used to identify the transaction period and liable to capture a part of the temporal dependency process. Such choices may affect the linear regression residuals and then spatial dependence detection and correction. However, in a spatial context, the DGP definition also implies considering spatial relations by integrating a spatial weights matrix. The specified models control for the presence of a binary matrix, the use of a threshold distance or the use of a nearest neighbour matrix. In the case of a continuous matrix, the type of distance function is also controlled.

Finally, in the case of an autoregressive specification, the estimation by the ordinary least squares method is no longer possible. Keeping the maximum likelihood (ML) estimation as a reference, the model also considers the two-stage least square estimation (TSLS) and the generalized method of moments (GM).

5.2.7 Descriptive statistics

Based on the collected information, three distinct data sets are created. The first is designed to explain the variations in spatial autocorrelation detection (Moran's I) and the two others consider the variations in spatial autocorrelation correction (ρ and λ). Each collected Moran's I statistic is measured on residuals from an ordinary least squares (OLS) estimation. Some references may be used in both analyses and each study can specify multiple econometric models with different specifications. Thus, several statistics can be collected by the study. The first uses 101 Moran's statistics (I) while the second and third respectively target the autoregressive coefficient and include 88 rho (ρ) coefficients and 213 lambda (λ) coefficients (Table 5.3).

Table 5.3: Sample of included studies

Sample 1 - Moran's I coefficient

Authors	N	%	Authors	N	%	Authors	N	%
BOWEN et al., 2001	3	2.97	HELBICH et al., 2013	1	0.99	NELSON, 2010	4	3.96
CAN, 1990	12	11.88	KESTENS et al., 2006	7	6.93	NOONAN, 2007	1	0.99
CAVAILHES et al., 2009	1	0.99	KRYVOBOKOV, 2010	6	5.94	OSLAND and PRYCE, 2012	5	4.95
COHEN et COUGHLIN, 2007	1	0.99	LI and SAPHORES, 2012	1	0.99	OSLAND et al., 2007	8	7.92
DEKKERS and Van Der STRAATEN, 2009	1	0.99	LOCHL et al., 2010	4	3.96	OSLAND, 2010	3	2.97
DES ROSIERS et al., 2010	1	0.99	MIHAESCU and VOM HOFE, 2012	1	0.99	SALVI, 2003	3	2.97
DES ROSIERS et al., 2011	5	4.95	MILITINO et al., 2004	5	4.95	SAMARASINGHE and SHARP, 2010	1	0.99
DUBE and LEGROS, 2014	20	19.8	NEELAWALA et al., 2013	1	0.99	SEDGLEY et al., 2008	2	1.98
HUI et al., 2007	1	0.99	NAPPI-CHOULET et al., 2009	3	2.97			
							101	100

Sample 2 - Lambda (λ)

	N	%		N	%		N	%
BOXALL et al., 2005	5	5.68	DUBE and LEGROS, 2014	7	7.95	NELSON, 2010	2	2.27
BIN et al., 2008	1	1.14	FERNANDEZ-ALVILES et al., 2012	2	2.27	NEWBURN et al., 2006	2	2.27
BRASINGTON and SAMARA, 2008	2	2.27	GEOGHEGAN et al., 2003	3	3.41	OSLAND and PRYCE, 2012	7	7.95
COHEN and COUGHLIN, 2007	4	4.55	HAM et al., 2013	2	2.27	OSLAND, 2010	3	3.41
CARRION-FLORES and IRWIN, 2010	3	3.41	HUI et al., 2007	1	1.14	PARENT and VOM HOFE, 2013	1	1.14
DEKKERS and Van Der STRAATEN, 2009	1	1.14	KRYVOBOKOV, 2010	5	5.68	PATERSON and BOYLE, 2002	3	3.41
DERMISI and McDONALD, 2010	1	1.14	KUETHE, 2012	1	1.14	PERDOMO, 2011	1	1.14
DES ROSIERS et al., 2010	1	1.14	LIAO and WANG, 2012	1	1.14	SAPHORES and LI, 2012	1	1.14
DES ROSIERS et al., 2011	3	3.41	LOCHL et al., 2010	1	1.14	TSUTSUMI and SEYA, 2009	2	2.27
DUBE et al., 2012	6	6.82	MIHAESCU and VOM HOFE, 2012	1	1.14	WILHELMSSON, 2002	7	7.95
DERMISI and McDONALD, 2011	1	1.14	NEELAWALA et al., 2013	1	1.14	YU et al., 2012	6	6.82
							88	100

Sample 3 - Rho (ρ)

	N	%		N	%		N	%
ANDERSSON et al., 2010	2	0.94	HELBICH et al., 2014	1	0.47	OSLAND and THORSEN, 2013	4	1.88
ARMSTRONG and RODRIGUEZ, 2006	3	1.41	HINDSLEY et al., 2013	3	1.41	OSLAND, 2010	3	1.41
ALBERTO et al., 2010	4	1.88	KIM and GOLDSMITH, 2009	9	4.23	PARENT and VOM HOFE, 2013	1	0.47
ANSELIN and LAZANO-GRACIA, 2008	6	2.82	KOSTOV, 2009	3	1.41	PATTON and McERLEAN, 2003	2	0.94
BELL and BOCKSTAEEL, 2005	8	3.76	KRYVOBOKOV, 2010	5	2.35	PRICE et al., 2010	3	1.41
BEAMONTE et al., 2008	2	0.94	KIM et al., 2003	3	1.41	PACE et al., 1998	1	0.47
BEAMONTE et al., 2010	3	1.41	LANDRY, 2011	9	4.23	PACE et al., 2000	3	1.41
BEAMONTE et al., 2013	1	0.47	LI and SAPHORES, 2012	1	0.47	PERDOMO, 2011	1	0.47
BRANDT and MEANNIG, 2011	6	2.82	LAZRAK et al., 2014	4	1.88	SALVI, 2003	2	0.94
BRANDT and MEANNIG, 2012	6	2.82	LEGUIZAMON and ROSS, 2012	6	2.82	SAMARASINGHE and SHARP, 2010	1	0.47
BRASINGTON, 2004	3	1.41	LIU, 2013	11	5.16	SANDER and HAIGHT, 2012	1	0.47
CAPOZZA et al., 2005	9	4.23	LOCHL et al., 2010	1	0.47	SEDGLEY et al., 2008	2	0.94
COHEN and COUGHLIN, 2007	3	1.41	MADDISON, 2009	1	0.47	SEYA and TSUTSUMI, 2013	3	1.41
CONWAY et al., 2010	1	0.47	MIHAESCU and VOM HOFE, 2012	1	0.47	SAPHORES and LI, 2012	1	0.47
CRESPO and GRET-REGAMEY, 2013	1	0.47	MILITINO et al., 2004	4	1.88	SMITH and WU, 2009	2	0.94
DERMISI and McDONALD, 2010	2	0.94	NEELAWALA et al., 2013	1	0.47	TSUTSUMI and SEYA, 2009	2	0.94
DERMISI and McDONALD, 2011	1	0.47	NAPPI-CHOULET et al., 2009	12	5.63	WALSH et al., 2011	3	1.41
FERNANDEZ-ALVILES et al., 2012	3	1.41	NAPPI-CHOULET et al., 2011	9	4.23	WILHELMSSON, 2002	1	0.47
HAM et al., 2013	2	0.94	NELSON, 2010	2	0.94	WALLNER, 2012	5	2.35
HANSZ and HAYUNGA, 2012	6	2.82	NOONAN, 2007	3	1.41	YU et al., 2012	6	2.82
HUI et al., 2007	1	0.47	OSLAND and PRYCE, 2012	2	0.94	ZHU et al., 2011	2	0.94

Table 5.4: Descriptive statistics - 3 samples: Moran's I ; λ and ρ

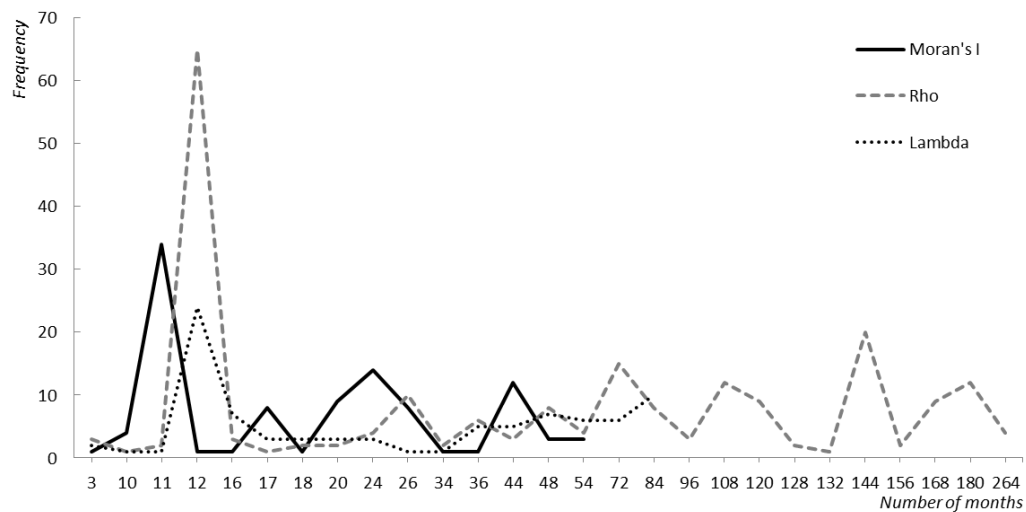
<i>Dependent variable: Moran's I, lambda, rho</i>	<i>Sample 1 - Moran's I</i>					<i>Sample 2 - Lambda (λ)</i>					<i>Sample 3 - Rho (ρ)</i>				
	<i>N</i>	<i>Mean</i>	<i>Std. Error</i>	<i>Min</i>	<i>Max</i>	<i>N</i>	<i>Mean</i>	<i>Std. Error</i>	<i>Min</i>	<i>Max</i>	<i>N</i>	<i>Mean</i>	<i>Std. Error</i>	<i>Min</i>	<i>Max</i>
Spatial statistic/coefficient	101	0.1631	0.1245	-0.015	0.563	88	0.4987	0.3100	-0.09	0.998	213	0.3907	0.2781	-0.126	0.983
<i>Independent variable: Study and dataset</i>															
Number of observations	101	6132	7034	193	25357	88	10155	14058	203	49594	213	16475	33319	64	190441
Number of observations (log)	101	7.928	1.397	5.263	10.141	88	8.180	1.583	5.313	10.812	213	8.083	1.903	4.159	12.157
Study location: North America (1/0)	101	0.525	0.502	0	1	88	0.602	0.492	0	1	213	0.488	0.501	0	1
Study location: Europe (1/0)	101	0.455	0.500	0	1	88	0.330	0.473	0	1	213	0.437	0.497	0	1
Study location: Other (1/0)	101	0.020	0.140	0	1	88	0.068	0.254	0	1	213	0.075	0.264	0	1
Study focus: methodological purpose (1/0)	101	0.594	0.494	0	1	88	0.307	0.464	0	1	213	0.385	0.488	0	1
Study focus: Case study (1/0)	101	0.406	0.494	0	1	88	0.523	0.502	0	1	213	0.451	0.499	0	1
Market: Office (1/0)	101	0.030	0.171	0	1	88	0.023	0.150	0	1	213	0.070	0.256	0	1
Market: Housing (1/0)	101	0.970	0.171	0	1	88	0.955	0.209	0	1	213	0.901	0.299	0	1
Real estate journals	101	0.535	0.501	0	1	88	0.273	0.448	0	1	213	0.211	0.409	0	1
<i>Independent variable: Dataset time range</i>															
Dataset temporal range (number of months)	101	65.31	52.16	10	180	88	58.48	49.29	3	144	213	70.69	64.04	3	264
Dataset temporal range (log of number of months)	101	3.727	1.050	2.303	5.193	88	3.597	1.059	1.099	4.970	213	3.729	1.115	1.099	5.576
<i>Independent variable: Data generating process</i>															
Presence of continuous distance variables (1/0)	101	0.347	0.478	0	1	88	0.625	0.487	0	1	213	0.545	0.499	0	1
Presence of buffer distance variables (1/0)	101	0.228	0.421	0	1	88	0.364	0.484	0	1	213	0.164	0.371	0	1
Independent variable: proximity variables	101	4.723	9.035	0	37	88	6.818	6.505	0	31	213	7.108	15.406	0	120
Independent variable: other	101	2.416	2.446	0	10	88	5.091	11.452	0	103	213	3.573	7.992	0	103
Functional form: Linear (1/0)	101	0.208	0.408	0	1	88	0.068	0.254	0	1	213	0.141	0.349	0	1
Functional form: semi-log (1/0)	101	0.178	0.385	0	1	88	0.352	0.480	0	1	213	0.399	0.491	0	1
Functional form: Log-log (1/0)	101	0.614	0.489	0	1	88	0.580	0.496	0	1	213	0.460	0.500	0	1
Estimation: GM	N/A	N/A	N/A	N/A	N/A	88	0.045	0.209	0	1	213	0.033	0.179	0	1
Estimation: TSLS	N/A	N/A	N/A	N/A	N/A	88	0.023	0.150	0	1	213	0.089	0.286	0	1
Estimation: ML	N/A	N/A	N/A	N/A	N/A	88	0.932	0.254	0	1	213	0.638	0.482	0	1
Model: SAC	N/A	N/A	N/A	N/A	N/A	88	0.148	0.357	0	1	213	0.061	0.240	0	1
Presence of temporal fixed effects (1/0)	101	0.564	0.498	0	1	88	0.420	0.496	0	1	213	0.404	0.492	0	1
Weights matrix: binary (1/0)	101	0.416	0.495	0	1	88	0.318	0.468	0	1	213	0.479	0.501	0	1
Weights matrix: number of neighbours (1/0)	101	0.356	0.481	0	1	88	0.580	0.496	0	1	213	0.535	0.500	0	1
Weights matrix: number of nearest neighbours	24	11.42	6.97	1	20	31	8.10	7.32	0	20	93	11.10	10.35	0	31
Weights matrix: with a threshold (1/0)	101	0.713	0.455	0	1	88	0.511	0.503	0	1	213	0.460	0.500	0	1
Weights matrix: threshold distance	49	2.39	2.24	0.2	9	38	1.84	1.49	0	4	69	1.32	1.53	0	10
Weights matrix: inverse distance (1/0)	101	0.347	0.478	0	1	88	0.500	0.503	0	1	213	0.446	0.498	0	1
Weights matrix: square-d inverse distance (1/0)	101	0.238	0.428	0	1	88	0.159	0.368	0	1	213	0.066	0.248	0	1
Weights matrix: spatio-temporal (1/0)	101	0.129	0.337	0	1	88	0.068	0.254	0	1	213	0.258	0.439	0	1

The Moran's I data set is composed of studies with, on average, 6,132 observations collected over 65 months. The vast majority of the analysis refers to the residential market mainly in Europe and North America. These studies are mostly made in a methodological perspective, i.e., that the residential values are used for an instrumental case study and do not focus on an empirical problematic. The second sample (λ coefficient) has a higher average number of observations (10,155) collected over an average of 60 months. The third sample (ρ coefficient) has an even higher number of transactions (average of 16,475) and the same applies to the mean temporal range of the dataset (70 months). In the latter two cases, the residential market is also the most represented, but the case studies are more numerous and mainly take place in North America.

Nearly 54% of the Moran's I indexes are collected from papers published in specialized real estate journals. This number respectively drops to 30% and 21% for the lambda and rho parameters.

Concerning, the Moran's I statistic calculation, the choice of the spatial weights matrix is mainly based on the definition of a threshold beyond which the spatial relationships are considered null. On the one hand, the average number of neighbours, when used, is 11 for the Moran's I statistic and, respectively, 8 and 11 for the λ and ρ parameters. On the other hand, the average threshold distance, when used, is 2.3 km for the Moran index, 1.84 for the λ parameter and 1.32 for the ρ parameter.

Finally, the continuous distance specification is mainly considered as an inverse distance function and a spatiotemporal matrix is involved in 13% of cases for the global autocorrelation detection against respectively 6% and 26% for autoregressive coefficients λ and ρ . For each sample, Figure 3 presents the frequency of models considered in function of its temporal range.



Note: The figure presents the frequency distribution of the used studies' temporal range for the three samples. As illustrated the most frequent temporal window is approximately 12 months long.

Figure 5.3: Frequency distribution of the datasets time ranges

5.2.8 Estimation results: Spatial dependence detection, Moran's I index

Based on the 101 Moran's statistics two OLS models are specified (Table 5.5, models 1 and 2). The second model differs from the first by integrating variables relative to the weights matrix specifications and temporal information about the datasets. A significant increase of the explanatory power (R^2), from 0.252 to 0.579, is then observed.

Table 5.5: Estimation results

Independent variables	<i>Moran's I index</i>		<i>Lambda parameter (λ)</i>		<i>Rho parameter (ρ)</i>	
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
	<i>Coefficient</i>	<i>Sig.</i>	<i>Coefficient</i>	<i>Sig.</i>	<i>Coefficient</i>	<i>Sig.</i>
Constant	-0.0763		0.3547		0.2924	
Number of observations (log)	0.0297 *		0.0428 **		0.9311 *	
Study location: North America (1/0)	-0.0154		0.0330		0.1932 *	
Study location: Other (1/0)	Reference		Reference		0.0498 ***	
Study focus: methodological purpose (1/0)	0.0312		0.0002		-0.1050 **	
Study focus: Case study (1/0)	Reference		Reference		-0.0992 *	
Presence of continuous distance variables (1/0)	-0.0494		-0.0289		0.0567	
Presence of buffer distance variables (1/0)	0.1278 ***		0.1133 *		Reference	
Independent variable: proximity variables	-0.0010		0.0009		-0.1715 ***	
Independent variable: other	-0.0103		-0.0123		-0.1674 ***	
Functional form: Linear (1/0)	-0.0122		-0.1101 **		0.0019	
Functional form: semi-log (1/0)	0.0750		0.0251		-0.0032 ***	
Functional form: Log-log (1/0)	Reference		Reference		0.0073	
Weights matrix: binary (1/0)			-0.0586		-0.0033	
Weights matrix: number of neighbors (1/0)			-0.0941 *		-0.0033	
Weights matrix: with a threshold (1/0)			-0.0425		0.0012	
Weights matrix: inverse distance (1/0)			-0.0488		-0.0033	
Weights matrix: squared inverse distance (1/0)					0.1861	
Weights matrix: spatio-temporal (1/0)			-0.0754		0.0719	
Dataset temporal range (number of months)			0.0039 **		-0.4760 **	
Dataset temporal range (log of number of months)			-0.1848 **		0.0037	
Presence of temporal fixed effects (1/0)			0.0359		-0.1408	
Dataset temporal range (number of month)*real estate journals			-0.0008		0.0129	
Dataset temporal range (log of number of month)*real estate journals			0.0551		-0.0773	
Presence of temporal fixed effects (1/0)*real estate journals			-0.1804 *		-0.0032 **	
Estimation: GM			0.0595		0.0356	
Estimation: TSLS			-0.5512 ***		0.0329	
Estimation: ML			Reference		0.2533 ***	
Model: SAC			-0.4346 ***		0.0557	
			Reference		Reference	
			-0.4329 ***		-0.0924	
					-0.1820 *	
N	101	101	88	88	213	213
R2	0.252	0.579	0.548	0.737	0.342	0.504
aic	-142.584	-180.478	-1.281	-24.800	-4.865	-41.032
bic	-113.817	-125.560	30.924	37.133	38.832	43.000
vif	1.65	13.82	1.72	12.77	1.45	7.84

legend: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

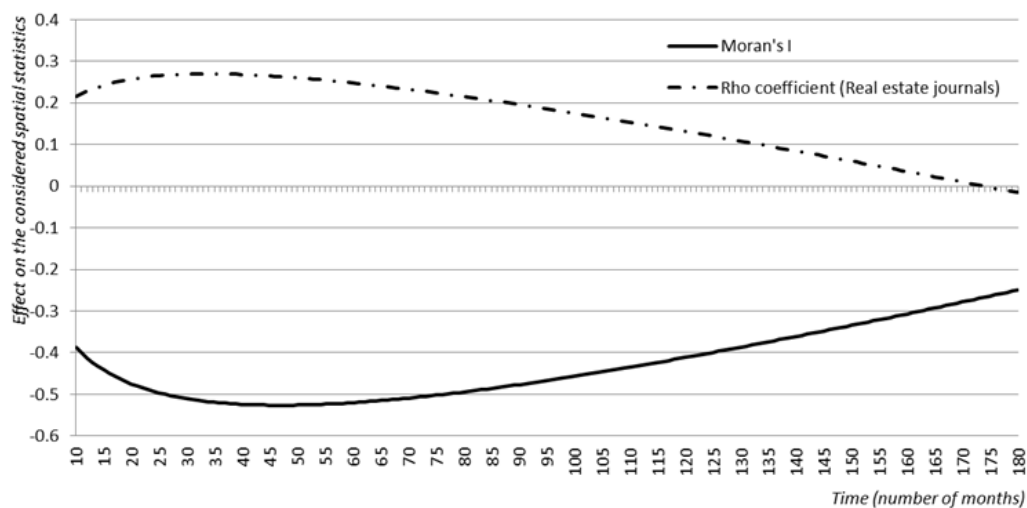
The sample size and the use of discrete spatial variables are significant in both models. The positive coefficients suggest that an increase in the dataset size results in a higher Moran's I index, which is in line with the over-connection problem exposed by Smith (2009). The variation of the Moran's I index does not seem significantly influenced by the type of spatial variables present in the specification. It may be difficult to distinguish the impact of continuous spatial variables (such as distance to given environmental amenities) and buffer spatial variables insofar as the model specification can include both. Moreover, the number of included spatial variables, without distinction (buffer or continuous), and the number of non-spatial independent variables does not significantly affect the measure of spatial autocorrelation. As spatial variables are used for environmental description, these results suggest that even if the variable omission is problematic, increasing the number of spatial variables does not necessarily mean a reduction of spatial autocorrelation. This encourages us to assign special importance to the choice and the quality of measure of these spatial variables.

Both models present a negative impact of the linear functional form of the hedonic pricing model (HPM), but the impact is only significant in the second model. This suggests that the HPM may well be highly non-linear. The linear functional form is then forcing the relation between the dependant and independent variables to be linear, which translates into the error terms and generates a spatial residual autocorrelation. It seems important to not consider spatial autocorrelation as a problem but as a data pattern that has to be optimally detected to be properly controlled. That said, even based on theory the *correct* DGP remains uncertain in HPM, but the choice of a given functional form necessarily implies some impact on the pattern of the residuals.

Considering the weights matrix specifications (model 2), it appears that only the use of the nearest neighbours' weights matrix specification significantly affects, negatively, the amplitude of the spatial autocorrelation statistic. This matrix specification forces spatial

relations between spatial observations to be limited to a determined number of neighbours. This result suggests that the over-connection problem may well lead to overestimation of the measured spatial autocorrelation. Interestingly, the use of a spatiotemporal weights matrix doesn't affect the autocorrelation detection statistic. This observation goes against previous studies' results that pointed to an overestimation of the Moran's statistic when time is omitted (Dubé and Legros, 2013a and 2013b). However, this may be explained by the different interpretation of temporal relations.

The highly significant coefficient associated with the number of months confirms the effect of the *time window* on spatial autocorrelation detection. Figure 4 illustrates the fluctuation of the effect depending on the temporal time range of the dataset.



Note: The figure illustrates the significant effect of the dataset temporal range on the Moran's I index and the lagged dependent variable coefficient (ρ). The global residuals' spatial autocorrelation increases and the transaction price autocorrelation decreases as a function of the dataset temporal length. This may be explained by different spatial distributions of dependency processes over time for estimation residuals and selling price trends.

Figure 5.4: Impact of the dataset temporal range on spatial statistics

As suggested by the signs of the coefficients, the curve associated with the Moran's I index (model 2 results) has a negative slope until an optimum (reached at 48 months), where it thereafter turns into a positive slope. Up to this optimum, the Moran's I statistic is, on average, lower than the mean value, but then increases. This result suggests that spatial relations are overstated as the temporal range of the dataset grows. The *density* of the dataset in terms of number of observations per unit of time may be a factor influencing the measurement of spatial dependence. In fact, spatial relations may be increased in a more compact dataset. This density should also be considered over space.

Finally, the inclusion of fixed effects controlling for the transaction period has distinctive effects for both groups of papers. In the case of the Moran's I , no significant impact on spatial autocorrelation measurement is found for unspecialized papers. In other words, temporal dummy variables do not appear to adequately capture effects of temporal dependence, which may affect spatial autocorrelation measurement as suggested by previous results. Adding fixed effect aims to capture (nominal) the price evolution of residential values. Moran's I indexes collected from papers published in real estate journals show a significantly lower spatial autocorrelation detection.

5.2.9 Estimation results: Spatial dependence correction

Regarding the autoregressive parameters, two similar models are specified depending on the choice of the DGP used in the papers. Both, exogenous specification (Table 5.5, models 3 and 4) and endogenous specification (Table 5.5 models 5 and 6) are presented.

Exogenous spatial process: SEM and lambda coefficient (λ)

Regarding the exogenous spatial process, the global performance of the first specification is quite interesting with an explanatory power (R^2) of 0.548. Adding variables referring to the weights matrix specifications and to the temporal dimension, the explanatory power (R^2) goes to 0.737 for the model 4.

According to the model 4, the λ parameter, referring to the exogenous spatial dependency processes, is significantly lower when continuous spatial variables are included in the model's specification. As suggested previously for the Moran's I , it appears that continuous variables may help to capture some kind of spatial relations, while not capturing the total effect. The functional form of the HPM has a significant impact on the amplitude of the spatial autoregressive coefficient. However, as shown in model 6, these effects disappear when information relative to the weights matrix specification and model estimation are added.

If the choice of the DGP does not affect the amplitude of the estimated coefficient, the choice of the estimation method plays a significant role. Compared to the maximum likelihood (ML) estimation method, the 2-stage least square (TSLS) method presents a strong negative impact on the λ parameter. Not surprisingly, when combined with an endogenous spatial lagged variable (ρ) the exogenous spatial lagged parameter is significantly lowered. In fact, the addition of the ρ parameter helps in capturing a part of the spatial dependence controlled by the λ coefficient. This suggests that a large part of residual autocorrelation may be explained by spatial and temporal patterns affecting property values. Consequently, a strong part of residual autocorrelation is due to the influence of neighbouring dependent variables values. This result reconsiders the way a specific DGP is chosen. While some statistical tests exist to discriminate between the two spatial DGP

based on Lagrange multiplier statistics (following a specific to general approach), there is still an estimation problem regarding a full specification that considers both spatial autoregressive effects in a unique DGP (Le Gallo, 2002).

Lastly, the parameter associated with the spatial lagged error term is greatly reduced when a spatiotemporal weights matrix is specified but no difference is noted with the inclusion on temporal fixed effects controlling for the transaction period. These results suggest that by increasing the dataset temporal range there is also an increase of the errors autocorrelation. The effect of time on the amplitude of the autoregressive parameter is thus only indirect here: it is related to the structure of the weights matrix accounting for the temporal dimension.

Endogenous spatial process: SAR and rho coefficient (ρ)

The ρ coefficient integrates endogenous spatial dependency processes. The model specification (Table 5.5, models 5 and 6) and data characteristics have a major impact on its value compared to the spatially lagged error coefficient. First, the sample size significantly increases the value of the estimated coefficient while studies using North American data present a significantly lower coefficient. Study context may affect the autoregressive specification. It is possible that, for a more spatially compact dataset, the property values tend to be closer and more related to each other, increasing endogenous spatial processes. By adding the weights matrix specification and temporal variables, the autoregressive variation explanation increases from 0.342 to 0.504.

The results show a negative relation between the presence of continuous spatial variables to define the environmental amenities. Moreover, the number of spatial variables negatively affects the ρ parameter even if the coefficient is weakly significant. Adding

continuous variables decreases the amplitude of the endogenous spatial autoregressive coefficient. With the inclusion of a larger number of variables describing local environment, a lower autoregressive coefficient is observed, suggesting that a more precise description of the neighbourhood captures some of the endogenous dependency process. In other words, those variables capture a portion of spatial underlying factors integrated in transaction prices and shared by nearest observations. Not surprisingly, the values tend to be similar in some places, not only because of their own characteristics, but also because of shared environmental amenities.

Concerning the model 6, the nearest neighbour's weights matrix specification increases the autoregressive coefficient. Because close residential values tend to be similar, by limiting relations to nearest neighbours it seems normal to observe a stronger and significant ρ coefficient. As for the λ parameter, the temporal range of the database does not directly affect the determination of the endogenous spatial dependence lagged parameter. However, observations collected in dedicated real estate journals are influenced by the dataset temporal range. As exposed in Figure 4, unlike the Moran's I index, the value of the estimated Rho coefficient tends to increase until a 38-month optimum before decreasing and returning to the reference value (the mean value over the number of studies at hand).

The inclusion of time fixed effects to the econometric specification helps to significantly reduce the lagged dependent variable coefficient for specialized real estate papers only. This result appears logical in the context of real estate values known to fluctuate over time. It also states that for this autoregressive process, the influence of time is mainly indirect: the amplitude of the coefficient being statistically related to inclusion of the temporal dummy variables. These observations highlight a specific effect for journals dedicated to real estate analysis.

5.2.10 Discussion

As outlined by Legendre (1993), spatial dependence is not a “problem” but has to be well managed since it may reveal something else. On the one hand, a spatially lagged dependent coefficient (ρ) appears to bias the other estimated coefficients (β) as well as their variance. On the other hand, a spatially lagged coefficient (λ) has no consequence on the estimated coefficients, but affects the precision of their estimated variance. Their fluctuations may generate bias in estimations and study conclusions.

According to the results of the MA, the measure and correction of spatial dependence is a complex process. Using information on spatial amenities is an effective and proper way to capture some part of the spatial processes. Results outlined the impact of discrete spatial variables on the Moran’s I index. Indeed, a binary definition of the environment increases the strength of spatial relations and the measure of spatial autocorrelation. However, the rho parameter is more sensitive to continuous distance and to the number of dependent spatial variables. The negative coefficients suggest that such types of variable adequately capture spatial process from the dependent variable autocorrelation.

Moreover, a linear functional form negatively affects the measured spatial autocorrelation but has no impact on any of the estimated spatial autoregressive coefficients. Interestingly, the weights matrix specification doesn’t explain much variation of the considered statistics. In fact, only a nearest neighbour weights matrix specification show significant results. It reduces the Moran’s I index but increases the rho coefficient. In both cases, the weights matrix is specified using an average of 11 neighbours but the dataset sizes are quite different, with an average of 6,132 observations for the Moran’s I index and 16,475 observations for the rho parameter. These statistics may reflect the effect of the dataset spatial density since a bigger sample may be more compact, meaning a

smaller mean distance between observations, which will therefore more likely be autocorrelated. The positive impact of the dataset size supports this assumption.

The rho parameter is more sensitive to the dataset characteristics and model specification than the lambda coefficient. Yet, the presence of temporal fixed effects in the model specification does not influence the autoregressive parameters. Interestingly, some particular effects are identified for observations collected from real estate papers such as the presence of dummy temporal variables and the dataset temporal range that significantly influence the dependent variable spatial autocorrelation coefficient (ρ). No distinction is observed for the residual spatial autocorrelation coefficient (λ). The non-significant result observed for the SEM specification outlines possible bias in such specifications and suggests that the SAR specification would be more appropriate.

The inclusion of both lambda and rho parameters in the DGP is used to estimate a full model (SAC). Both coefficients appear to be lowered when such a model specification is applied. It is difficult to determine which coefficient affects the other since the true empirical data structure remains unknown. However, as noted by Piras and Prucha (2014), if the true DGP excludes one of these lagged parameters, the use of a full model specification creates a loss of efficiency.

This real estate journals effect may outline a distinctive review approach and could suggest different model or weights matrix specification practices between specialized and non-specialized journals. The effect may also be due to different dataset characteristics. There are 15,000 sells on average for observations collected from non-real estate journals, this number increases to 22,135 for specialized journals. A simple mean comparison test is performed to compare both groups but no significant distinction can be made. The same exercise is performed using the dataset temporal ranges. Results show that samples used in analyses published in real estate papers are significantly longer (respectively 62 and 102 months on average). According to a variance comparison test, no significant difference is

made between both groups. Thus, results tend to confirm that a longer dataset range affects the lagged dependent variable spatial autoregressive coefficient, while a series of temporal fixed effects dummy variables adequately captures market fluctuations of nominal price evolution.

Finally, the spatiotemporal specification of the weights matrix strongly and negatively affects the lagged error term (λ) but not the other spatial statistics. This suggests that errors are likely to be temporally dependent but such a process differs from that of the sell prices. In fact, the latter's dependency structure can be attributed to market fluctuations which may be more homogeneously spatially distributed and well captured by temporal fixed effects (real estate papers). On the other hand, it is likely that error terms' autocorrelation is fluctuating differently over time from place to place as a result of localized exogenous events affecting the real estate market.

5.2.11 Conclusion

Considering the consequences of spatial autocorrelation on estimation results, the research aimed to define the influence of the temporal dimension on the detection and correction of spatial autocorrelation. A meta-analysis approach has been proposed to overcome the limitations of previous research mainly based on single samples with fixed characteristics.

Three hypotheses were defined. The first hypothesis assumes that time significantly influences spatial statistics but that the magnitude of such impact depends on the model data generating process (DGP). Globally, the first research assumption is partly confirmed. The meta-analysis reveals that temporal dependency does affect the amplitude of the Moran's I index and, in some cases, the autoregressive coefficients as well. The λ

coefficients are lower when the temporal dimension is considered in the weights matrix specification.

In fact, the inclusion of fixed effect controlling for the period of transaction affects the ρ coefficient as well as the Moran's I index. The data generating process (DGP) also affects spatial statistics depending on the type and number of included spatial variables. This is particularly true for the residual and autocorrelation detection statistic and for the endogenous spatial dependence processes, while the lagged error term is less sensitive.

The second hypothesis proposes that the elapsed time between the first and the last observation affects calculated spatial statistics and estimated spatial autoregressive coefficients. According to our results, the temporal span of the dataset significantly influences the autocorrelation detection statistics but not the lagged error parameters. For its part, the dependent variable autocorrelation is also influenced by the temporal range but only for papers published in real estate journals. The dataset's temporal and spatial density may explain these results. Thus, the second hypothesis is partly confirmed and outlines a possible over-connectivity problem (Smith, 2009), which should then not be limited to spatial relations but also applied to temporal proximity.

Finally, the third hypothesis suggests that characteristics of the studies influence spatial autocorrelation detection and correction. The conclusions outline that the number of observations especially influence the lagged dependent variable coefficients and the autocorrelation detection index. This supports the role of the dataset density considering that the closer the properties are in time and/or space, the more likely they are to be similar.

Thus, multiplying the introduction of environmental variables does not fully solve the spatial autoregressive pattern hidden through residuals. An equilibrium between the data characteristics and the formalization of their structure through a model specification is necessary but difficult to implement because of the lack of recommendations. Overall, the

conclusions outlined a distinct behaviour of autocorrelation processes depending on the lagged variable type. However, as mentioned by Le Sage and Pace (2014a), the influence of the spatial weights matrix specification is limited. The authors suggest that the *myth* of the spatial weights matrix may be explained by past work based on misspecified weights matrices and models.

Spatial methods sometimes seem to be applied mechanically and are similar from one study to another, which offers little variability in the approaches and therefore in the test variables to include in such meta-analysis. It is then difficult to more precisely consider the type of spatial variable, an optimal distance radius of influence for spatial relations. However, the research provides a good overview of the consequences of the most used approaches in the detection and correction of spatial autocorrelation for disaggregated data in a spatiotemporal context. Further research should verify if such results also apply to specification tests used to distinguish both processes such as a Lagrange multiplier. Unfortunately the literature used in this meta-analysis does not report enough specification tests yet, which makes it difficult to extend the actual analysis to such a case.

CHAPITRE 6

QUAND TOBLER RENCONTRE HAGERSTRAAND : L'IMPORTANCE DES DIMENSIONS SPATIALES ET TEMPORELLE DANS L'ANALYSE STATISTIQUE

6.1 PRÉSENTATION DE L'ARTICLE

Le second article de la thèse, intitulé « When Tobler meets Hägerstraand : About the importance of spatial and temporal dimensions in statistical analysis », a été rédigé en collaboration avec les professeurs Jean Dubé et Diego Legros. La méthodologie proposée a notamment fait l'objet d'une présentation sur invitation à l'Université de Bourgogne (Dijon, France) au département d'économie de Dijon (LEDi) en 2014. La recherche a également mené à une présentation dans le cadre du congrès annuel de l'*American Association of Geographers* (AAG) à Chicago en 2015 et a été subventionnée par le Fonds de recherche sur la société et la culture (FQRSC).

L'article a été soumis à la revue *Letters in Spatial and Resource Sciences* et s'inscrit dans le prolongement direct de la première méta-analyse présentée au premier article de la thèse et vise à répondre aux limites rencontrées par cette première analyse.

La première MA consistait à dresser une collection d'études empiriques permettant d'évaluer l'impact de l'omission de la dimension temporelle sur les statistiques spatiales.

Une des limites importantes de cette première analyse repose sur le fait qu'elle ne permet de considérer efficacement que les méthodes les plus utilisées dans la littérature, alors qu'elle échoue à intégrer les approches alternatives récentes. Pour ce second article, une approche différente est utilisée : la méthodologie est inspirée de la MA, alors que les cas d'espèces sont construits à partir d'une base de données portant sur les transactions immobilières enregistrées dans Lucas County, OH (É.U.) entre 1993 et 1998.³¹ La construction d'un nombre important de sous-échantillons permet ainsi d'obtenir les statistiques d'intérêt (I de Moran et paramètres autorégressifs) en fonction de diverses spécifications du DGP et de la matrice de pondérations. Cette stratégie est également inspirée d'une approche de type *bootstrap* permettant de tester l'impact de l'omission de la dépendance temporelle et le rôle de la matrice de pondérations (spatiales vs spatiotemporelles) sur les statistiques d'intérêt.

Les résultats suggèrent que le coefficient autorégressif des modèles SAR, ρ , est plus sensible aux caractéristiques de l'échantillon et aux choix de modélisation lorsqu'une approche strictement spatiale est appliquée. En conséquence, les effets spatiaux risquent d'être mal interprétés et les conclusions des études, d'en souffrir. Pour sa part, l'indice I de Moran apparaît moins affecté par ces facteurs.

³¹ Cette base de données est disponible gratuitement sur le site internet de James LeSage ainsi que dans la librairie de MatLab. Elle a également été utilisée par les travaux de Dubé et Legros (2013).

6.2 WHEN TOBLER MEETS HÄGERSTRAAND: ABOUT THE IMPORTANCE OF SPATIAL AND TEMPORAL DIMENSIONS IN STATISTICAL ANALYSIS

6.2.1 Abstract

Modeling a spatiotemporal process structuring spatial micro data pooled over time, such as property values, is a challenge. While spatial autocorrelation may generate estimation bias invalidating study results, some empirical case studies ignore the temporal dimension assuming a simultaneous collection of data, and potentially overestimation spatial relations. Previous studies outlined the necessity to opt for a spatiotemporal approach but relied on small sample analysis or simulated data. An original methodology inspired from meta-analysis (MA) and relying on empirical data is proposed to assess the impact of the omission of the temporal dimension on spatial autocorrelation detection statistics (Moran's I) and on an estimated spatial autoregressive coefficient (ρ). The results confirm that spatial statistics are more sensitive to sample characteristics and exogenous modeling choices, with a strictly spatial approach producing a larger autoregressive coefficient.

6.2.2 Introduction

Tobler's first law of geography states that "Everything is related to everything else, but closer things more so" (1979: 236). This affirmation is somewhat similar to what is understood as temporal autocorrelation, with the temporal persistence of values and shocks over some time periods. Indeed, as mentioned by Hägerstraand (1970), the spatial pattern cannot be separated from their temporal constraints. Spatial relations between observations are not synonymous of causality (Miller, 2004). The tendency to add the temporal

dimension to the analysis represents a major step in the understanding of spatial and regional phenomena (Arbia, 2011, Anselin, 2002) due to their dynamic nature and thus, to a certain extent, the relevance of the analysis beyond a descriptive logic. In fact, from a strictly spatial point of view, causality mechanisms remain almost impossible to establish.

That being said, spatiotemporal models are mostly developed in the perspective of panel or pseudo-panels (Elhorst, 2014a). The application of spatial econometric modeling tools to spatial micro data pooled over time (unrepeatedly collected data), such as real estate values, is one of the relevant challenges a spatial analyst faces: “since individual housing sales do not occur at regular time intervals, it is difficult to model such processes in terms of standard discrete time series” (Smith and Wu, 2009: 334). By nature, spatial micro data pooled over time are not necessarily repeated, while their aggregation into pseudo-panel implies a loss of information due to aggregation problems as well as clearly generating the Modifiable Areal Unit Problem (MAUP, see Openshaw 1984).

The paper aims at analyzing the influence of omitting the temporal dimension in the modeling process of spatial micro data collected over time. Using real estate data, a methodology inspired from meta-analysis (MA) is proposed: i) to assess the impact of a strictly spatial analysis on spatial statistics, controlling for the effects of the sample characteristics and ii) to overcome limitations of previous work on the topic whose conclusions generally value a spatiotemporal approach. An empirical case study is proposed using 25,357 real estate transactions occurring in Lucas County (Ohio, USA) between 1993 and 1998. Results based on different subsamples of different sizes and temporal lengths show that ignoring the temporal constraints on the spatial relations can lead to high divergence in estimated spatial autoregressive coefficients and thus, to overestimation of the spatial spillover effect.

The paper is divided into six sections. The first section proposes a brief review of literature dealing with spatiotemporal modeling based on spatial micro data pooled over

time. The second section discusses the methodology based on meta-analysis and the associated meta-regression framework. The third section presents a methodological framework related to the meta-analysis and how such a framework inspired the methodological framework using a simple data set. The fourth section describes the data used for the empirical case study and the sub-sample scheme used to build distinct databases, while the fifth section focuses on estimation results. A brief conclusion closes the paper.

6.2.3 Spatiotemporal data and hedonic price modeling: a brief literature review

Inspired by Tobler's "first law of geography" outlining the relations between spatially distributed elements (Tobler, 1979), spatial econometric tools are developed to consider the spatial distribution of subjects of interest and to deal with spatial dependence structuring them. The presence of spatial dependence in the data is not a problem strictly speaking but needs to be controlled to ensure reliable results since the standard ordinary least squares (OLS) estimators are likely to lose their property and efficiency if regression residuals display spatial patterns (Elhorst, 2001). Spatial autocorrelation is a central concept in empirical spatial research (Getis, 2008), but econometric tools developed are applied with little consideration for the challenges posed by it or its effects over time (Ye and Rey, 2013).

The consideration of spatial effects in econometric models is motivated by the theory as well as the specifics of the data used by empirical studies. The models with spatial interactions therefore manage the links between economic agents that could lead to collective behavior or aggregation patterns that emphasize the role of location (Anselin, 2002), with spatial dependence usually captured using an exogenously specified weights matrix. Even if the role of the spatial weights matrix is discussed (LeSage and Pace, 2014a,

Anselin, 2002), it remains the cornerstone of spatial econometrics methods (Bhattacharjee et al., 2012, Chen, 2012, Elhorst, 2010). Its specification is an unavoidable challenge for empirical research since it affects autoregressive coefficients (LeSage and Pace, 2009a) and therefore the measure of the total, direct or induced effects.

So far, spatial statistics depend upon the weights matrix specification (Dubé and Legros, 2014a, Dubé and Legros, 2013b, Dubé and Legros, 2013a). Spatiotemporal modeling is mainly applied for panel and pseudo panel analysis (Elhorst, 2014a) using observations repeatedly measured over time. Since, all data may not be continuously observed, the challenge lies in applying these methods to spatial micro data pooled over time (e.g., individuals, firms, property values, etc.) respecting the multidirectional and unidirectional nature of spatial and temporal dependency processes (Dubé and Legros, 2013b, Dubé and Legros, 2013a).

Some suggest treating the information as a pseudo-panel, i.e., by aggregating the observations based on their location and the moment of their observation (Huang et al., 2006). This strategy helps in formalizing spatiotemporal dependency processes but also raises analytical problems such as a loss of information due to aggregation (Haining et al., 2009, Anselin, 2002, Anselin and Griffith, 1988) and the *Modifiable Areal Unit Problem* (MAUP, see Openshaw, 1984; Dubé and Brunelle, 2014). Consequently, many real estate hedonic analyses still use a strictly spatial approach to explain property values in a wide variety of problematics (Devaux and Dube, 2016). The consideration of the spatial spillover effect changes depending on the inclusion of the temporal dimension affecting study results, since the autoregressive parameter is used to compute the *Average Total Impact* (ATI).³²

32 For details on the computation of the Average Total Impact (ATI) see LeSage and Pace or Drukker et al. .

Among the proposed strategies, two spatiotemporal perspectives stand out. The first is the *Spatiotemporal Autoregressive model* (STAR) proposed by Pace et al. (1998, 2000). The basis of this approach is to introduce a *filter* to allow isolating the spatial, temporal and spatio-temporal effects through a decomposition based on spatial and temporal weights matrices. According to Liu (2013), such a spatiotemporal approach can significantly reduce prediction errors and has been widely applied in empirical or methodological analysis (Tu et al., 2004, Nappi-Choulet and Maury, 2011, Nappi-Choulet, 2009, Beamonte et al., 2010b, Beamonte et al., 2010a, Beamonte et al., 2008).

The second spatiotemporal model suggests combining both matrices, spatial and temporal, using an Hadamard product³³ to adequately isolate multidirectional and unidirectional spatial effects (Dubé and Legros, 2013a, Smith and Wu, 2009, Huang et al., 2010). This approach is sometimes designated as the *Spatiotemporal Hedonic Model* (STHM, see Hui et al., 2015). However, the proposed strategy is not restricted to real estate analysis but is applicable to a wide variety of themes using microdata pooled over time. Unlike the STAR model, the specification of a general spatiotemporal weights matrix allows estimating autoregressive parameters for interpreting spatial dynamic effects.

According to Dubé and Legros (2013a, 2013b) this spatiotemporal modeling approach is appropriate to isolate/decompose the spatial autocorrelation according to temporal constraints on spatial relations. The authors clearly underline the importance of the temporal dimension in the price determination process, which can otherwise be attributed to the spatial dimension: “Considering the spatial dimension alone when data have an important temporal dimension can lead to the overestimation of spatial dependence

33 The Hadamard product consists of a term-by-term matrix multiplication: $\mathbf{W}_{(N_T \times N_T)} = \mathbf{S}_{(N_T \times N_T)} \odot \mathbf{T}_{(N_T \times N_T)}$, where N_T is the total number of observations, taking into account the fact that total sample size is different in each time period t .

in residuals and, consequently, can lead to overcorrection using the appropriate spatial econometric approaches” (Dubé & Legros, 2013a: 8). For their part, Thanos et al. (2016: 87) “are calling for a systematic approach to spatiotemporal data analysis in applied economics, formalizing this currently ‘fragmented’ field.”

Based on the second spatiotemporal modeling perspective, the paper aims to: i) verify the impact of the omission of the temporal dimension when using spatial econometric procedures for spatial micro data pooled over time, such as real estate transactions; ii) assess the impact of the sample characteristics on spatial statistics and iii) compare strictly spatial and spatiotemporal modeling approaches. To achieve these objectives, a research strategy based on MA principles is proposed.

6.2.4 Methodology

As described by Glass (1976 in Florax et al., 2002 :27): “Meta-analysis refers to the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings.” The MA is relevant for synthesizing information and knowledge or theory building in a context of divergent study results and is a valuable approach in the regional or urban analysis context (See for example: Debrezion et al., 2007, Melo et al., 2009, Nelson, 2004, Schipper et al., 2002, Sirmans et al., 2010).

Research in a controlled environment is generally not possible in social sciences and *Monte Carlo* experiments may have difficulty in reproducing the complexity of spatiotemporal dependencies structuring the observations. In consequence, the MA is an interesting approach for comparing modeling strategies. However, an MA generally does not control for the applied methodology from which estimation variation may occur (Pigott, 2012). Despite the abundance of case studies, a classic literature review may imply a

considerable rejection rate of potential studies to include in an MA due to missing information or imprecise methodology description (Devaux and Dube, 2016). Moreover, the MA realized in a methodological perspective and based on a literature review may be limited to the “most used” approaches and fail to compare more recent or innovative strategies. The analysis would then be limited to the most common modeling habits.

The MA generally aims at estimating a mean effect (*Effect Size*, ES) for a collection of empirical studies and to explain the observed variation in this effect from one study to another. The ES (equation 6.1) corresponds to the mean effect weighted by the inverse of its variance (Borenstein et al., 2009):

$$ES = \frac{\sum_{s=1}^S P_s \pi_s}{\sum_{s=1}^S P_s} \quad (6.1)$$

Where, π_s is the considered effect for a sub-sample s , with $s = 1, 2, \dots, S$, and P_s is the associated weight defined by the inverse of its variance: $1/\sigma_{\pi_s}$. For the present analysis, two distinctive statistics (effects) are considered and represented by the Moran’s I index and the spatial autoregressive parameter, ρ .

Beyond calculation of a mean effect, the MA also aims at explaining the fluctuation of this effect between the different sub-samples considered by means of regression models (or meta-regression – MR) (Harbord and Higgins, 2008a). The MR model is used to extract and explain the proportion of between sample variations, ψ^2 , from the total variation of the considered statistic (Higgins and Thompson, 2002a).³⁴ According to Borenstein et al.

³⁴ The authors normally refer to the between study variance as the I^2 statistic. It is renamed ψ^2 in this study to distinguish it from a Moran’s I index.

(2009) and as shown in Figure 6.1³⁵, the ψ^2 statistic can be understood as the measure of inconsistency between study results (spatial statistics in this case). It is used in the present study to assess the fluctuations of the spatial statistics between the different sub-samples.

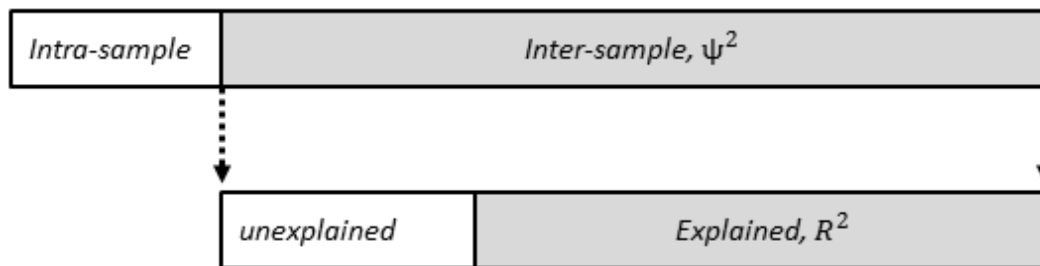


Figure 6.1: Variation explained by the meta-regression model

The MR model aims at explaining this measured effect. The random effect MR model captures possible heterogeneity in the residuals and assumes that the effect variable follows a normal distribution. Equation 6.2 presents the model as proposed by Harbord and Higgins (2008a).

$$\mathbf{Y} = \mathbf{M}\boldsymbol{\theta} + \mathbf{W}\boldsymbol{\varphi} + \mathbf{u} + \boldsymbol{\delta} \quad (6.2)$$

Where \mathbf{Y} is a vector of spatial statistic of interest (Moran's I index or ρ parameter), of dimension ($S \times 1$); \mathbf{M} is a matrix of independent (or explanatory) variables related to the data generating process (DGP) of dimension ($S \times K$), \mathbf{W} is a matrix related to the specification of the weights matrix used to obtain the spatial statistics, of dimension ($S \times$

³⁵ Adapted from Borenstein et al. 2009: 202.

V). The vector of parameters θ and φ are, respectively, of dimension $(K \times 1)$ and $(V \times 1)$ and isolate the impact of the DGP specifying the hedonic price model and the specification of the weights matrix on the obtained spatial statistics. Finally, the two remaining vectors, u and δ , are of dimension $(S \times 1)$ and allow to isolate, respectively, the random effect element capturing heterogeneity among residuals, assuming a zero mean and a τ^2 variance³⁶, and a white noise term. The random effect, δ , aims at capturing possible heterogeneity among the unexplained between study variance (Harbord and Higgins, 2008a).

6.2.5 Data and Sampling Procedure

The strategy relies on the construction of sub-samples extracted from a large dataset. Each sub-sample randomly varies in terms of size and temporal range (the time between the first and last observation). The full dataset is composed of 25,357 single-family property transactions observed in Lucas County (Ohio, USA) between 1993 and 1998.³⁷

The sampling procedure consists of a four-stage process. First, a total of 100 subsamples are randomly collected from the full database using random temporal and sample sizes.³⁸ Second, spatial statistics are computed and estimated from each of the subsamples using different model data generating processes (DGP, model and weights matrices specifications) as shown in Figure 6.2. Third, standard regression using an ordinary least squares (OLS) hedonic price model is estimated. For each sample, the OLS models are estimated with and without monthly temporal fixed effects controlling for market trends. Such fixed effects could help in capturing a part of the temporal dependency

36 For more details on the estimation of τ^2 , the interested reader can refer to Pigott or Hedges and Vevea .

³⁷ The dataset is publicly available in the MatLab library for spatial econometrics.

³⁸ The maximum sample size is limited to 10,000 to speed up calculation time of the needed spatial statistics.

processes even in a strictly spatial modeling approach.³⁹ Since the functional form can affect hedonic modeling estimations, both linear and log linear functional forms are estimated for comparison purposes. The residuals are used to compute Moran's I indexes and variances. Fourth, the same subsamples are used to estimate spatial autoregressive (SAR) models using a different specification of the weights matrices: i) one that does not take into account the temporal constraints of spatial relations; and ii) another one to isolate the spatial multidirectional relations. For each of these two types of matrices, three different transformations, based on distance among the observations (d_{ij}), are used to build spatial weights: i) one that considers the inverse distance function ($1/d_{ij}$); ii) one that uses the negative exponential transformation ($e^{-d_{ij}}$); and iii) another one that uses an indicator variable ($d_{ij} \leq d_i^*$), for a total of 6 different weights matrices per sub-sample selected.⁴⁰ The estimated autoregressive coefficients and their respective variance are conserved.

³⁹ Monthly fixed effects are used considering the fact that: i) this is the smallest time range of some of the created samples and ii) they are widely used in empirical literature.

⁴⁰ The mean distance between observations is defined as the maximum threshold distance beyond which the spatial weights matrix's elements are composed of null values.

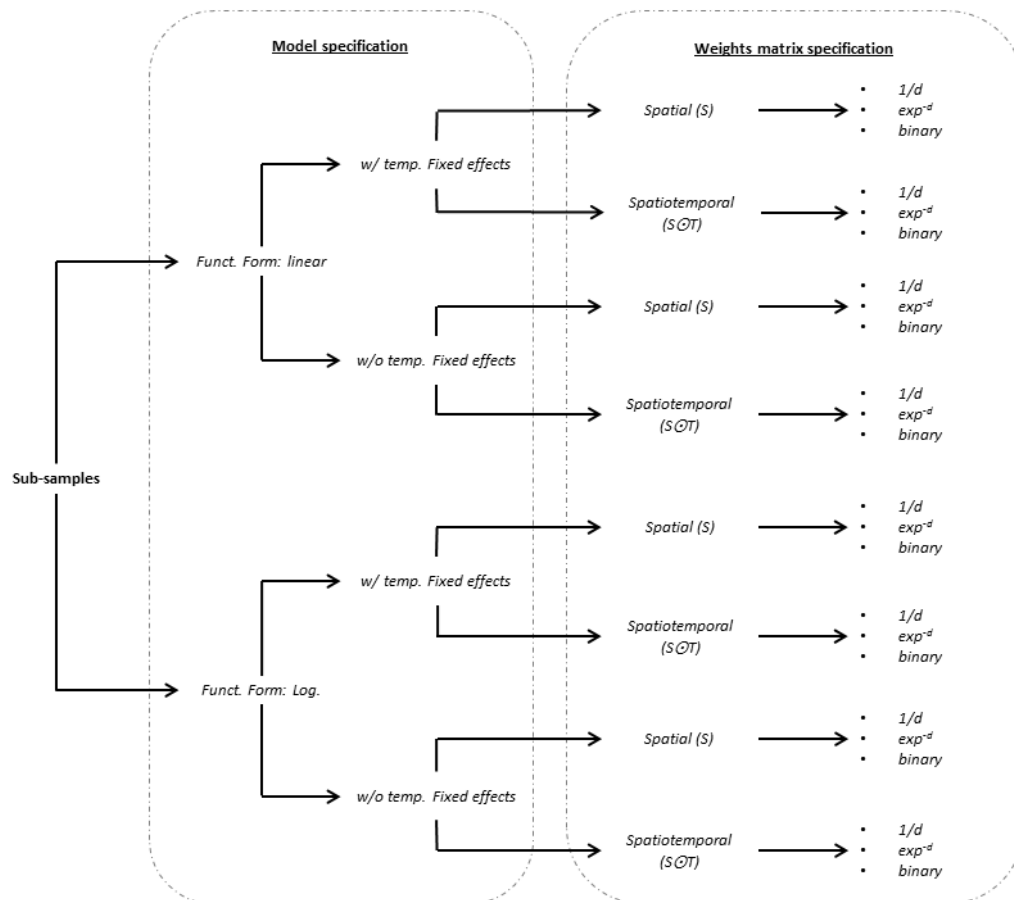


Figure 6.2: Meta-analysis dataset building process

This sampling and calculations process leads to obtaining 2,400 spatial statistics (Moran's I indices and spatial autoregressive coefficient, ρ). However, since the absence of neighbors could lead to computation problems, all samples including at least one observation without neighbors are rejected. The final dataset is composed of 1,632 Moran's

I indices and autoregressive coefficients (ρ)⁴¹ of which one half (816 observations) is estimated using a spatial weights matrix and the other half (816 observations) using a spatiotemporal weights matrix.

The Moran's *I* statistic varies between -0.0217 and 0.2925 and its variance oscillates between 0.0003 and 0.0133. The ρ parameter fluctuates between -0.0011 and 0.7065. The mean subsample size for individual estimation is 4,828 transactions distributed over a 36-month time range. The mean distance between the observations is 10.5 km and the mean threshold distance is 6.7 km. There is an average of 1,359 neighbors for each transaction when a strictly spatial matrix is used, while this number drops to 397 for the temporally restricted matrix (Table 6.1).

⁴¹ The use of a standardized weights matrix produces an autoregressive coefficient and Moran's *I* indexes bounded between -1 and 1.

Table 6.1: Descriptive statistics

<i>Variables</i>	Descriptive statistics			
	<i>N</i>	<i>Average</i>	<i>Min</i>	<i>Max</i>
<i>Moran's I index</i>				
Moran's I index (Spatial + spatiotemporal)	1632	0.1059	-0.0217	0.2925
Moran's I index variance (Spatial + spatiotemporal)	1632	0.0029	0.0003	0.0133
Moran's I index (Spatial)	816	0.1042	-0.0156	0.2694
Moran's I index variance (Spatial)	816	0.0021	0.0003	0.0075
Moran's I index (Spatiotemporal)	816	0.1076	-0.0217	0.2925
Moran's I index variance (Spatiotemporal)	816	0.0036	0.0007	0.0133
<i>Autoregressive coefficient (SAR model)</i>				
Autoregressive coefficient, ρ (Spatial + spatiotemporal)	1632	0.3844	-0.0011	0.7065
Estimated coefficient standard error (Spatial + spatiotemporal)	1632	0.0236	0.0040	0.0648
Autoregressive coefficient, ρ (Spatial)	816	0.4428	-0.0011	0.7065
Estimated coefficient standard error (Spatial)	816	0.0246	0.0100	0.0648
Autoregressive coefficient, ρ (Spatiotemporal)	816	0.3261	-0.0011	0.6126
Estimated coefficient standard error (Spatiotemporal)	816	0.0226	0.0040	0.0593
<i>The samples</i>				
Sample size	1632	4827.6	1033	9393
Dataset time range (months)	1632	35.9	4	69
Density of observations per period of time (N/months)	1632	168.5	22.1	509.7
<i>The models</i>				
Number of explanatory variables	1632	27.0	9	78
Temporal fixed effects, month of sale (1: yes; 0: no)	1632	0.5	0	1
Functional form (1: log.; 0: linear)	1632	0.5	0	1
<i>The weights matrices</i>				
Weights matrix type	1632	3.5	1	6
W: strictly spatial (1: yes; 0: no)	1632	0.5	0	1
W: spatiotemporal (1: yes; 0: no)	1632	0.5	0	1
Mean distance between observations (km)	1632	10.5	10.0	10.9
Threshold distance (km)	1632	6.7	2.4	10.5
Maximum distance between observations	1632	58.0	52.6	59.5
Average number of neighbors, spatial W matrix	1632	1359.0	144.3	4430.5
Average number of neighbors, spatiotemporal W matrix	1632	397.1	23.8	1103.5
W: negative exponential distance function (1: yes; 0: no)	1632	0	0	0
W: inverse distance function (1: yes; 0: no)	1632	0.3	0	1
W: binary matrix based on the threshold distance (1: yes; 0: no)	1632	0.3	0	1

The spatial statistics estimated are used as dependent variables in an MR while the information concerning the DGP, weights matrix specification and sample characteristics act as regressors. Figures 6.3a to 6.3d present the spatial statistics' distributions.

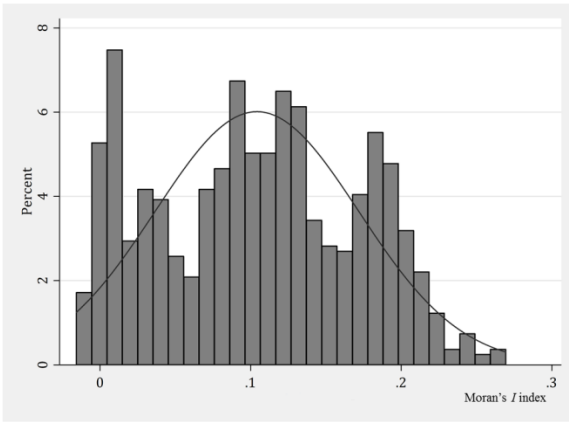


Fig.6.3a - Moran's I Index, spatial matrix

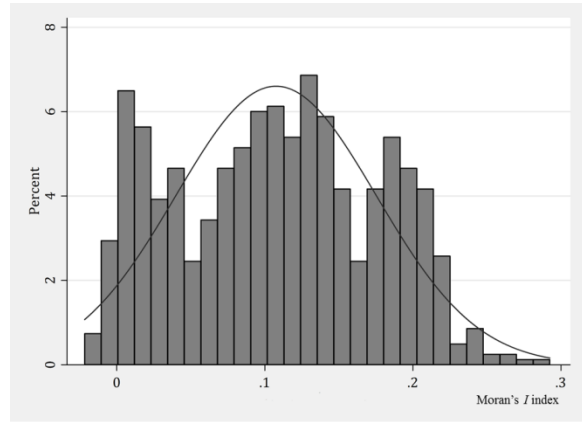


Fig.6.3b - Moran's I Index, spatiotemporal matrix

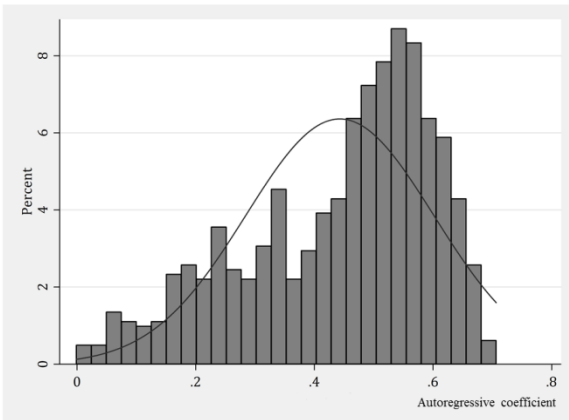


Fig.6.3c - Rho coefficient, spatial matrix

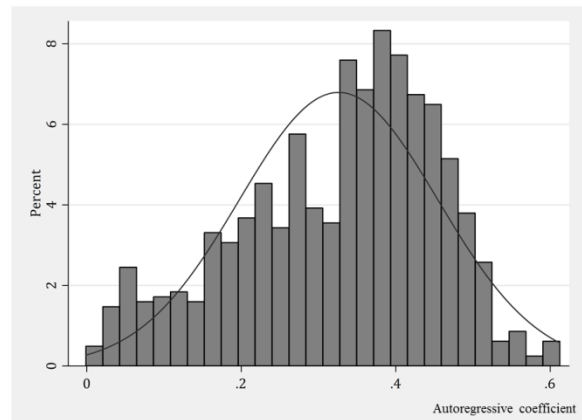


Fig.6.3d - Rho coefficient, spatiotemporal matrix

Figure 6.3: Spatial statistics' distributions

6.2.6 Analysis results

Mean effect size

Concerning the spatial autocorrelation detection (Moran's I index), all mean effects show significant evidence of spatial autocorrelation (Figure 6.4). Since all lines are superposed in the figure, no significant distinction is noted regarding the presence, or not, of temporal fixed effects regardless of the matrix type. Similar confidence intervals are also observed for the two types of matrices. Unlike results found by Dubé and Legros (2013b) omitting the temporal dimension doesn't seem to affect the risk of finding significant evidence of spatial dependence. Interestingly, the use of temporal dummy variables with a spatiotemporal matrix minimizes the Moran's I index, suggesting that a major part of the spatial dependence captured is actually due to lack of control for the temporal dimension. Overall, a spatiotemporal weights matrix specification does not seem to affect the spatial dependence detection as much as suggested by an empirical MA performed by Devaux and Dubé (2016).

When modeling spatial dependence, the importance of correctly specifying the weights matrix is more affected than detection (Figure 6.5). The model based on a full spatial weights matrix, with no regard for temporal constraint, returns a significantly higher value of the autoregressive parameter as compared with the approach restricting the spatial relations within the same time period. In both cases, the inclusion of temporal fixed effects do not significantly affect the coefficients. Again, the variance of the autoregressive term is slightly increased in the spatial perspective. Overall, the mean effect results suggest that the omission of the temporal dimension for an SAR specification may overestimate spillover effects and affect the other regression estimators.

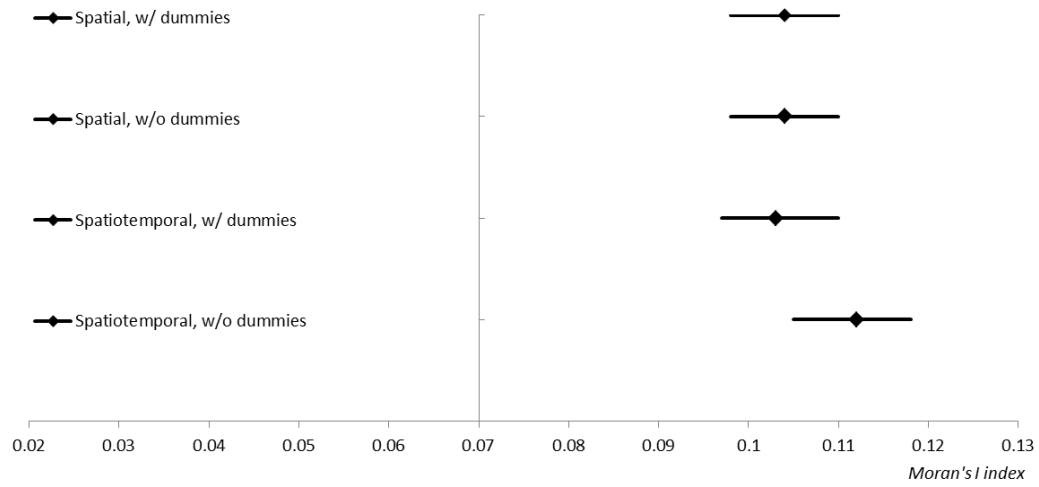


Figure 6.4: Comparison of the mean effects size for the Moran's I index

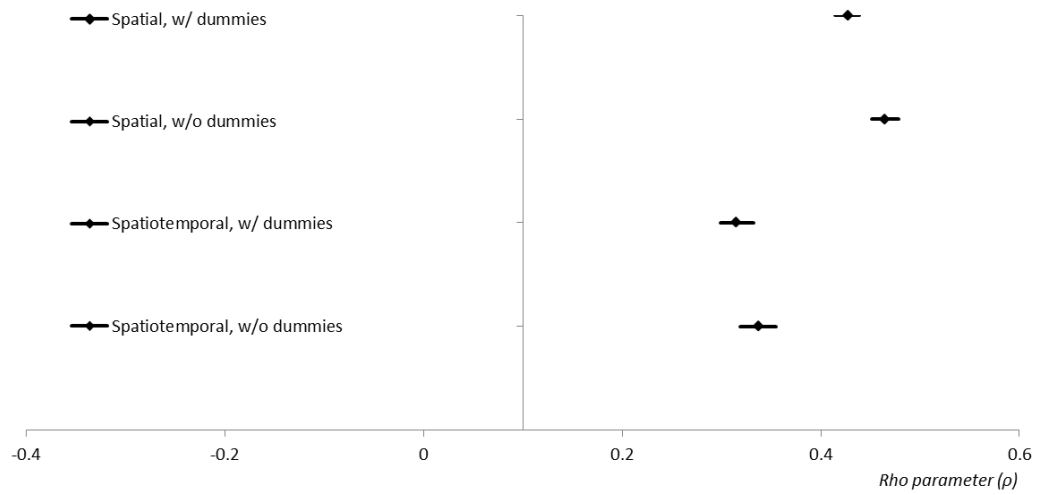


Figure 6.5: Comparison of the mean effects size for the Rho (ρ) coefficient

Meta-regression results:

The MR models are performed to explain the variation between the considered spatial statistics. For each statistic, three series of models are estimated (Table 6.2). The first (model a) includes results for all specifications combined (spatial and spatiotemporal) while the two others (models b and c) use the results based respectively on spatial specifications and spatiotemporal specifications.

According to the ψ^2 statistic, representing the explained variation between the subsamples, no major distinction is observed between the Moran's I index calculated using a spatial or spatiotemporal weights matrix. This suggests that the differences in the Moran's I index are due to the sub-samples' characteristics and the estimation choices but not to exogenous elements. The small τ^2 statistic, associated with the variance of the random element of the regression model, confirms this observation.

The similar R^2 statistics (models 1a, 1b and 1c) reveal that the Moran's I index is similarly affected by the estimation approaches (model and weights matrix specification) and the sample characteristics independent from the weights matrix used. Model 1a allows comparing both specification approaches. In line with the ES results, the non-significant coefficient associated with the spatial weights matrix specification confirms that the omission of the temporal does not impact on the spatial autocorrelation detection.

Table 6.2: Estimation results

Variables	Moran's I index [†]						Spatial autoregressive model (SAR) [†]					
	Model 1a		Model 1b		Model 1c		Model 2a		Model 2b		Model 2c	
	S+ST		S		ST		S+ST		S		ST	
	coef.	sig.	coef.	sig.	coef.	sig.	coef.	sig.	coef.	sig.	coef.	sig.
Constant	0.18347	***	0.19044	***	0.16017	***	0.40834	***	0.59169	***	0.29065	***
Density of observations per period of time (N/months)	8.53E-05	***	9.25E-05	***	7.59E-05	***	6.90E-06		-3.74E-05		3.77E-05	
Dataset time range (months)	4.47E-04	***	5.37E-04	***	4.06E-04	***	-3.42E-04		4.44E-04		-1.12E-03	*
Sample size	-1.26E-06	*	-2.38E-06	**	1.62E-06		1.19E-05	***	8.90E-06	***	2.45E-05	***
Average number of neighbors	1.87E-06		1.16E-05		-6.11E-06	*	1.59E-05	*	2.38E-05		-1.78E-05	
Functional form, linear (1: yes; 0: no)	Reference		Reference		Reference		Reference		Reference		Reference	
Functional form, logarithm (1: yes; 0: no)	0.03989	***	0.04003	***	0.03974	***	0.00445		0.06446	***	-0.05363	***
Number of explanatory variables	1.76E-04	*	-6.53E-06		3.64E-04	***	2.38E-04		3.12E-04	*	1.61E-04	
Absence of temporal fixed effects, months (1: yes; 0: no)	Reference		Reference		Reference		Reference		Reference		Reference	
Temporal fixed effects (1: yes; 0: no)	0.00232		-0.00019		0.00492		-0.01310		-0.01619		-0.01050	
Threshold distance (km)	-0.01084	***	-0.01173	***	-0.00819	***	-0.01231	***	-0.01957	***	0.00493	
W: negative exponential distance function (1: yes; 0: no)	Reference		Reference		Reference		Reference		Reference		Reference	
W: inverse distance function (1: yes; 0: no)	-0.04312	***	-0.04125	***	-0.04505	***	-0.02501	***	-0.04414	***	-0.00521	
W: binary matrix based on the threshold distance (1: yes; 0: no)	-0.12451	***	-0.12427	***	-0.12478	***	-0.17447	***	-0.27272	***	-0.07597	***
Spatiotemporal weights matrix (1: yes; 0: no)	Reference						Reference					
Spatial weights matrix (1: yes; 0: no)	-0.00164						0.13459	***				
N	1632		816		816		1632		816		816	
τ^2	0.0007		0.0007		0.0007		0.0121		0.0048		0.0124	
ψ^2	0.9953		0.9970		0.9877		0.9878		0.9149		0.9887	
Adjust. R²	0.8453		0.8451		0.8512		0.4888		0.7976		0.2259	
Prob > F	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000	

Legend: * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; †: All observations have at least 1 neighbor;

S: Spatial weights matrix;

ST: Spatiotemporal weights matrix, 20 days before and 10 days after transaction;

Comparing model 1b (spatial weights matrices) and model 1c (spatiotemporal weights matrices), most of the variables appear to have similar impacts. In both specifications, the density of observation increases the measure of autocorrelation detections as well as the dataset time range (Table 6.2). Since the spatial weights matrices limit spatial relations between observations independently of the moment of their realization, this suggests that an over-connection of the spatial relations leads to an increased spatial autocorrelation detection identification (Smith, 2009). However, an increased threshold distance decreases the Moran's I index, suggesting that more distant observations may be less related and then less autocorrelated (Table 6.2). This observation is coherent with the well-known rationale of Tobler's first law of geography. The results associated with the type of distance function used to build spatial weights confirm this point since a binary matrix, imposing similar weights regardless of proximity, reduces the Moran's I index.

Some other distinctions between results for models 1b and 1c should be noted though. The sample size significantly reduces the Moran's I index measured using a strictly spatial definition when building a spatial weights matrix, which is not the case when the spatial relations are limited within a given time period, i.e., spatiotemporal weights matrices (Table 6.2). Inversely, the mean number of neighbors has a negative effect on the spatiotemporal matrices only, which implies that an increased number of relations decreases the Moran's I index (Table 6.3). While the number of neighbors increases, the Moran's I decreases. It should also be mentioned that controlling for the temporal nominal aspect of sale prices using a temporal fixed effect does not depict an influence on spatial autocorrelation detection, as measured by the Moran's I index.

Comparing model 2b and 2c, the between subsamples explained variance (ψ^2) statistic is lower using the strictly spatial weights matrix, suggesting a slightly larger influence of external factors. Unlike the Moran's I index, the autoregressive coefficient is

significantly increased using the strictly spatial weights matrix. The omission of significant spatial variables is known to affect spatial dependence and then spatial statistics. It is plausible that a strictly spatial approach increases this problem by creating spatial relations between observations influenced by similar spatial variables and then overestimating the mean spillover effect (ρ).

According to the R^2 statistics, a greater share of the between sample variance is explained in model 2b. This observation is confirmed with an increased τ^2 term (random element variance), suggesting a more important effect of exogenous factors' variation. Thus, the spatial specification explains a large part of the autoregressive coefficient, suggesting an overestimation of the spatial effect, as previously mentioned by Smith (2009). Among the different model specifications affecting results in a strictly spatial context, the weights matrix distance threshold and the use of an inverse distance function (compared to a negative exponential distance function) reduce the autoregressive parameter but have no significant effect in a spatiotemporal context. Such observations may be explained by the higher weights granted to distant observations when the inverse distance function is used. The farthest observation could be less similar and contribute to minimizing the measured average spillover effect (ρ). Moreover, the negative impact of the threshold distance in a strictly spatial specification implies a decreasing ρ coefficient while the distance increases. This observation also encourages this assumption of dissimilarities between distant observations. The fact that similar results are not observed in a spatiotemporal specification supposes that such an approach efficiently limits the impact of exogenous specification choices on study results.

Even if the sample size is positively associated with the autoregressive parameter in both specifications, the effect is the most important when time is considered. In a spatial context, a larger dataset should imply an increased number of interactions between observations. But again, in a strictly spatial context, it seems that more relations also mean

potentially more connections between less similar values and consequently a smaller, but still positively significant, average spatial dependence correction parameter. In other words, if connections between observations increase the ρ parameter, it seems that spatially and temporally distant observations tend to be less similar.

A wider dataset time range contributes to decrease the spillover effect measured using a spatiotemporal specification. However, using a strictly spatial specification, the time elapsed between the first and last observation doesn't affect the ρ coefficient. It is plausible that the number of observations is a more relevant source of variation than the dataset temporal range in a spatial perspective.

The use of temporal fixed effects has no impact on the spatial autocorrelation correction. One can assume that temporal price fluctuations more globally affect the collected transaction prices and that an autoregressive specification mostly captures localized effects.

Overall, the spatial specification leads to a more important measure of the average spillover effect that appears to be more related to a model and weights matrix specification. For its part, a spatiotemporal approach leaves more space for autoregressive term fluctuations due to exogenous factors.

6.2.7 Discussion and Conclusion

The process of modeling spatial (micro) data pooled over time is a challenge for scholars working with individual data. Recent development suggests some interesting avenues in how to deal with such data, while the impact of neglecting the temporal dimension is still under-documented. In this paper, we propose an MA based on the study of the variation of results within a given and large database. Different sample sizes based

on different temporal ranges are used to calculate spatial autocorrelation detection, based on the Moran's I index, as well as a spatial autoregressive coefficient based on two distinct constructions of weights matrices: i) one that does not take into account the temporal constraints on spatial relations (strictly spatial weights matrix); and ii) one that explicitly limits the spatial relations to the same time period (spatio-temporal weights matrix). Using data on single-family transactions occurring in Lucas County (Ohio) between 1993 and 1997, the empirical analysis reveals that the explicit consideration of the temporal dimension through the use of a spatio-temporal weights matrix reduces the autocorrelation coefficient ($\hat{\rho}$) where a binary threshold distance spatial matrix is used.

Results thus confirm the over-connection problem, mentioned by Smith (2009), and this is also in line with previous MA results based on real estate literature (Devaux and Dube, 2016). The explicit omission of the time dimension creates spurious relationships between observations that go against the temporal logic (*arrow of time*, Thanos et al., 2016). Consequently, a strictly spatial analysis is not suitable to assess a dynamic and causality process and could generate bias in spatial statistics at the expense of the validity of study results.

The paper outlines, once again, the necessity of considering the temporal dimension since it affects the estimation of spatial spillover effects. Such effects are widely used in the understanding of regional and urban dynamics.

CHAPITRE 7

IMPACT DES POLITIQUES DE PATRIMONIALISATION SUR LES VALEURS RÉSIDENIELLES : LE CAS DU VIEUX-QUÉBEC

7.1 PRÉSENTATION DE L'ARTICLE

Après deux méta-analyses, la thèse se poursuit par deux travaux de nature appliquée. Le premier article, intitulé « The Impact of Heritage Policies on Residential Property Values: The Case of the Old city of Québec », est une étude de cas réalisée en collaboration avec les professeurs Jean Dubé et Etienne Berthold. L'article a fait l'objet d'une présentation au congrès annuel de l'*American real estate association* (ARES) de 2015 à Fort Myers, FL (É.U.) et a été soumis à la revue *Urban Geography*.

L'étude vise à évaluer l'impact d'une politique de patrimonialisation sur les valeurs foncières dans le quartier historique du Vieux-Québec. En tant que partie intégrante d'une politique de patrimonialisation, le projet de développement urbain vise à entretenir l'héritage culturel intangible, mais est également susceptible d'engendrer des impacts économiques, notamment par le biais de la rente de localisation et, par conséquent, via le marché immobilier. L'approche méthodologique est donc inspirée de la théorie hédonique et de l'équation de prix hédonique (première étape).

Les travaux scientifiques tentent généralement d'estimer l'effet de la localisation dans un district historique (Clark and Herrin, 1997; Leichenko et al., 2001 ; Narwold et al., 2008). Pourtant, au-delà de la désignation historique d'un secteur, un ensemble de

politiques ciblées contribuent au maintien de l'héritage culturel et à la symbolique des lieux. À notre connaissance, aucun article n'a jusqu'ici tenté d'évaluer l'impact d'une politique de patrimonialisation par le marché immobilier.

Plus précisément, la démarche de l'article repose sur l'application de la méthode des ventes répétées et sur l'estimation d'un modèle de différence-en-différences (DID) (Gibbons and Machin, 2008). Le modèle DID a aussi été vanté pour sa capacité à réduire les problèmes de spécification associés au choix de la forme fonctionnelle (McMillen, 2010). L'utilisation de la différence première permet d'isoler les effets d'un changement exogène sur les valeurs foncières (la restauration de la rue d'Auteuil) tout en contrôlant les effets des aménités intrinsèques et extrinsèques constantes dans le temps. De ce fait, cette stratégie méthodologique permet d'ignorer les effets de certains schémas spatiaux notamment associés à la présence de variables omises (Dubé et al., 2014). En d'autres termes, en contrôlant les effets spatiaux, l'estimateur DID permet de cibler certains changements dans le temps.

À des fins de comparaison, la proximité à la rue est considérée en utilisant la distance euclidienne et la distance réseau. Les résultats suggèrent que les travaux d'aménagement urbain n'ont pas eu d'effet significatif sur les valeurs foncières à proximité immédiate de la rue d'Auteuil. En revanche, un effet négatif est observé autour de la zone ciblée (entre 150 et 300 mètres lorsque la distance euclidienne est considérée et entre 150 et 450 mètres lorsque la distance réseau est utilisée).

L'effet implicite sur les prix de vente est considéré comme étant la différence entre les externalités négatives et positives (externalités nettes). Ainsi, l'effet positif associé à la dimension symbolique et historique du projet pourrait être annulé par certains choix relatifs à l'aménagement physique de la rue. Celle-ci a en effet été rétrécie, les trottoirs ont été élargis, mais surtout le revêtement de sol asphalté a été remplacé par des pavés, générant ainsi un bruit plus intense lors du passage de voitures. Ces choix d'aménagement urbain

pourraient avoir eu un impact sur le flux de circulation et les nuisances sonores associées et avoir ainsi engendré des externalités négatives supérieures aux externalités positives.

Pour leur part, les rues voisines pourraient en revanche ne pas profiter des effets positifs et ne subir que les effets négatifs relatifs aux modifications physiques de la rue (déviation du trafic, stationnement difficile, nuisance sonore).

7.2 THE IMPACT OF HERITAGE POLICIES ON RESIDENTIAL PROPERTY VALUES: THE CASE OF THE OLD CITY OF QUÉBEC

7.2.1 Abstract

Few studies have attempted to measure the impacts of historic designation on residential property values through hedonic modeling. Most of the time, investigations have been limited to the premium associated with the fact of being localized inside a historic district. Yet not enough attention has been paid to the impacts of heritage policies *within* the historic district. This paper is intended to fill the gap by investigating the impact of such policies on residential properties for the case of the Old City of Quebec (Canada), designated by law as a protected area and included on the World Heritage List (UNESCO). More precisely, it aims to measure the impacts of the reorganization and landscaping of one particular street located south-west of the protection area on condominium house prices. A difference-in-difference (DID) estimator based on a hedonic repeated sales approach is proposed to isolate such effect based on proximity to the site under study. Estimation results suggest that the reorganization of the street had no significant impact on closest properties' prices, but had negative effects for properties located within 150 to 450 meters from the street. These observations outline complex relations between heritage policies and urban landscaping projects.

7.2.2 Introduction

As an important expression of contemporary identities, cultural heritage now pervades the world of urban studies. Urban regeneration projects focusing on restoration and historic building development are now at the forefront. A clear search for a form of meaning, heritage itself, is increasingly being approached as a social construction whose investigation necessarily involves ideologies and underlying socio-economic contexts (Berthold, 2012b, Graham et al., 2000, Hartog, 2014, Lowenthal, 2015, Pouliot, 2006). Policies are an essential part of heritage conservation. They give an impulse to conservation and find their concrete implementation in public action and governance (Graham et al., 2000, Guillaume, 1980). As is the case for many cultural policies (Saint-Pierre, 2008), heritage policies do not follow a static list of objectives. They reflect historic contexts in which they have been created and may change according to the evolution of societies. In fact, urban heritage policies should be distinguished from a standard landscaping policy since they carry intangible and symbolic goals. What has been less clear so far is whether such policies also translate into economic impact.

There are at least two complementary ways to approach the economic impacts of heritage policies. On the one hand, heritage is frequently regarded from the viewpoint of tourism benefits. From this perspective, the emphasis is placed on the multiplicative (direct, indirect and induced) economic benefits of tourism on local communities or on the fiscal impacts (Grefe, 2003). On the other hand, some attempts have been made to measure the economic impact of historic designation (Mason, 2005) through real estate values (Clark and Herrin, 1997, Leichenko et al., 2001, Narwold et al., 2008). However, the challenge remains to distinguish between the impact of a given historic policy and the fact of being located within a historic district. Moreover, such studies often disregard the other forms of urban preservation, policies or programs (Ryberg-Webster and Kinahan, 2014).

This paper investigates the impacts of a particular heritage policy on residential (condominiums) values in a historic district. The designation process is not sufficient to sustain cultural and ideological symbolism of the area and has to be underpinned by punctual heritage policies. The analysis is based on the case of the Old City of Quebec (Canada), the oldest part of the oldest city in Canada. Using hedonic pricing modeling, the research aims at assessing the effect of a particular heritage policy, as a part of a wider set of intervention in the context of an urban regeneration project focusing on historic building restoration. A difference-in-difference (DID) estimator based on a repeated sales approach (Dubé et al., 2011b, Dubé et al., 2013b) is applied to assess the impact of an exogenous landscaping policy (the d'Auteuil Street restoration) using a sample of 744 condominium multiple transactions within the old part of the city between 1997 and 2011.

In the dense context of a historic district, the definition of proximity can affect study results. Considering that the estimated effect may vary over space, Euclidean distance and the network distance are used to compare the local impact of proximity of historical landscape policy on residential real estate values. In both cases, no significant impact is observed for the properties located close to the site. However, results suggest that the reorganization of the street had a negative and significant effect on sale prices for properties located a little further from the street. These conclusions outline possible complex relations between the heritage policy and urban landscaping projects.

The paper is divided into five sections. The first part highlights the role of cultural heritage and heritage policies and presents the Old City of Québec and the case under study: the restoration of d'Auteuil Street. The second part attempts to summarize and discuss research that has addressed the impacts of historic designation on real estate values. The third section formally presents the methodological approach used in the current paper and presents the difference-in-difference estimator. The fourth section presents data used to

estimate the model while the fifth section discusses the main findings. Finally, the sixth section proposes a brief discussion and a conclusion.

7.2.3 The old city of Québec historic district and the reorganization of D'Auteuil Street as heritage policies

Cultural heritage and heritage policies: some theoretical principles

Cultural heritage may be defined as an object, a collection or an intangible heritage that has been judged worthy to be preserved, passed on and put on display (Berthold, 2012b). Basically, heritage is clearly a search for a form of meaning. The post-structuralist approach which is the basic paradigm for the concept sees heritage as part of a system defined by its capacity to create signs, a system upheld by institutions, standards, practices and values and symbols (Morisset, 2009).

Cultural heritage can also be defined in political terms, often in ways that seek to legitimize power (Ranger and Hobsbawm, 1983). As such, studying the construction of heritage necessarily involves ideologies and underlying socio-economic contexts (Rizzo and Mignosa, 2013, Throsby, 2010); in the world of urban planning, heritage is inseparable from an economic perspective. The emergence of cultural heritage is historically localized and cannot be isolated from the division of knowledge into separate disciplines that are a feature of modern societies. From this perspective, the study of heritage construction is also the offspring of prior knowledge constructs.

Public landscaping policies may play an important role in heritage conservation. On a concrete basis, they help implement heritage in public action and governance. Furthermore, they provide straight guidelines to urban planners. Heritage policies reflect the historic

contexts in which they were created. As dynamic measures, they may convey the political ideologies of a group seeking domination over another (Graham et al., 2000), support the commodification process of heritage to different extents (Zukin, 1996) or feed knowledge constructs (Berthold, 2012b).

In other words, urban planning in a historic district takes on a new meaning and dimension, playing an important role in heritage construction using symbolism and ideologies. In all cases, heritage policies must be seen as generating externalities possibly internalized through the real estate market, implying that it may be possible to recover its implicit price through a hedonic pricing model. Most of the time, property values are positively influenced by landscaping, generating a form of profitability of the urban development project due to tax income increases. However, to our knowledge, no prior empirical study has attempted to assess the economic effect of a particular heritage policy on property values in a historic district context. The Old City of Quebec is at the heart of important heritage policies, as the city of Quebec itself is one of the ancient cities in North America (founded in 1608). Thus the historic district is particularly well adapted to measure the possible economic effects of heritage policies.

The Old City of Quebec and the reorganization of D'Auteuil Street

The Old City of Quebec has been designated a historic district since 1963. It comprises 3.5 km² stretching from Champlain Street to St. Vallier St. East in the Lower Town, and the walled city as well as Côte d'Abraham in the Upper Town (see Figure 7.1). According to the 2011 census, its population was 4,780 inhabitants with important socio-economic disparities recorded among its different sectors. The most famous part of the Old City is clearly the *intra muros* section, which is more oriented towards touristic activities.

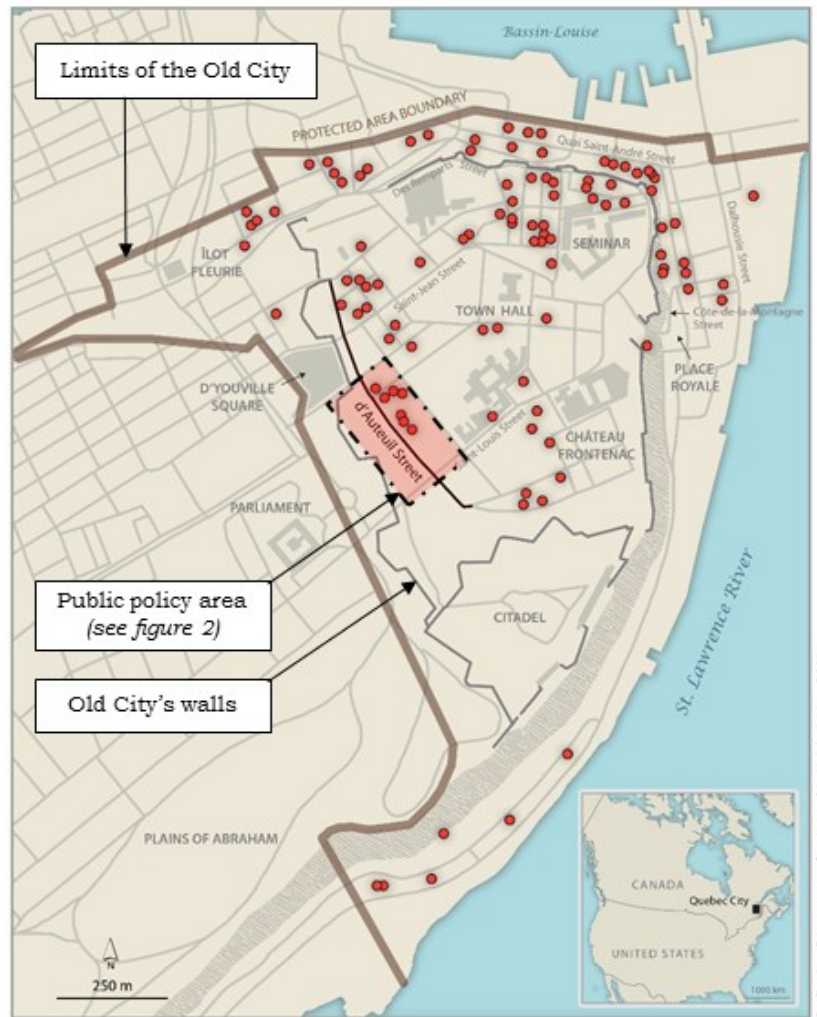


Figure 7.1: Spatial distribution of the repeated sales over the Old Quebec City

According to the Office du Tourisme, the number of persons that visit the site was estimated at 4,013,000 in 2008 (Berthold, 2012a). Tourism in the Old City has risen since 1985 when UNESCO listed the city of Quebec as a world heritage site. The Old city is also

an important central business district (CBD), with more than 15,000 people entering the district to work every day. The traffic flow was around 30,000 vehicles per weekday, most of them in transit (i.e. they did not originate in the Old City and were headed elsewhere). Thus, this part of the city regroups two major economic vocations, while also hosting a residential vocation.

Until the 1980s, the most decisive heritage policy in the Old City was the restoration of Place Royale, located outside the walls and right under the famous Château Frontenac. This state-run operation was marked by many excesses and led to the reconstruction of many houses and the reorganization of the urban fabric in order to present the Old City of Quebec as the birthplace of New France (Berthold, 2012b).

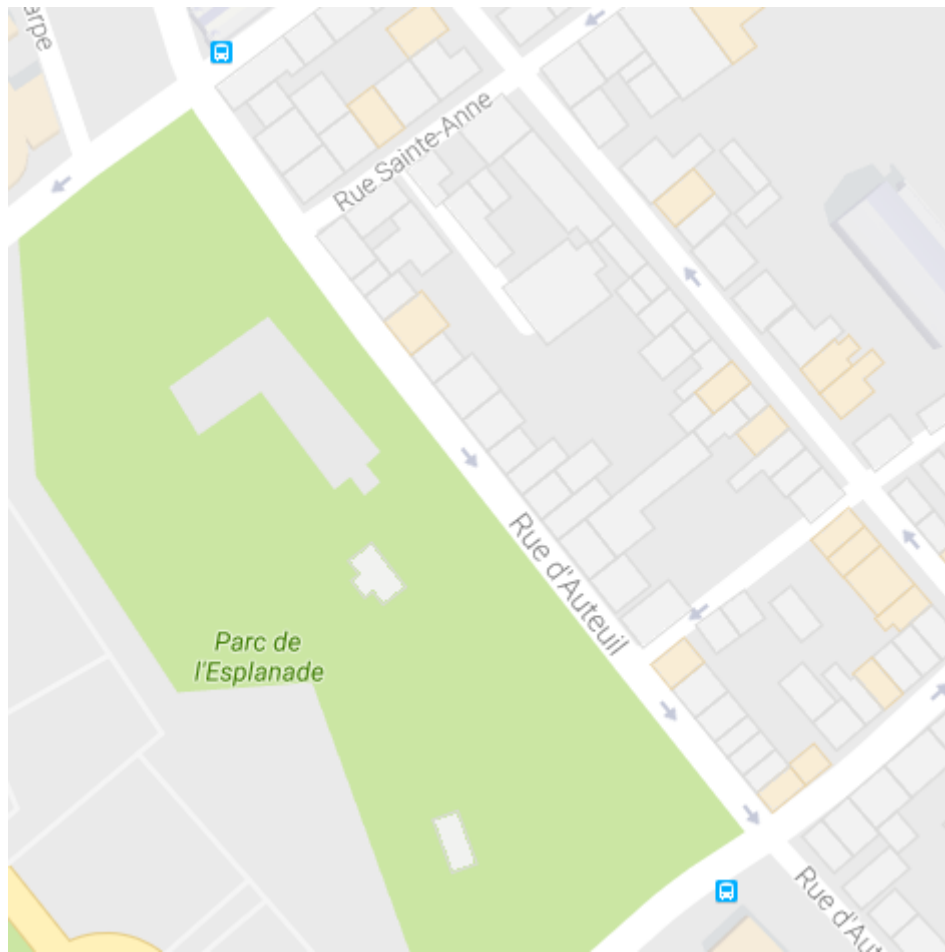
In 1979, due to criticism targeting the Place Royale project, the City of Quebec created an Old City and heritage division in its urban planning department. The creation of this division outlines the uniqueness of the Old City: a distinctive district with distinctive urban planning expectations. Under a formula that exists to this day, the division was given the mandate to manage the activities of public and private agencies in the Old City. In this regard, an important program, “Maître d’oeuvre,” was established in the 1980s. Still in operation in 2015 within the perimeter of the Old City, this program tried to stimulate residential restoration by offering to pay up to 40% of the cost of major repairs to the exteriors of buildings. The program granted a total of \$10.3M (CAD) between 1995 and 2010, greatly contributing to the residential restoration practices in the Old City. This ‘new’ and more integrative model marked a turning point to a more contemporary facet of heritage policies in the Old City.

By the end of the 1990s to the present day, historic commemoration and cultural heritage enhancement in the Old City of Quebec historic district has been partly conducted through the Commission de la capitale nationale du Québec (CCNQ). This organization was created in the mid-1990s by the Parti Québécois nationalistic government. The very

existence of such an organization highlights the peculiarities of historic districts in terms of urban development.

The CCNQ was basically given a threefold mandate: i) to reorganize the parliament hill, on which sits the Quebec National Assembly; ii) to promote Quebec national history through the historic commemoration of great men and celebrities; and iii) to promote internationally the city of Quebec as the Province of Quebec's national capital city, though in fact Canada's capital city is Ottawa. One way to do so was the commemoration of monarchs and other great men from all around the world. The reorganization and landscaping of d'Auteuil Street was deeply influenced by such heritage policy.

D'Auteuil Street is one of a hundred streets that can be found inside the walls of the Old City of Quebec. The street is located south-west of the historic district (see Figures 7.1 and 7.2). The East side of the street is composed of residential and commercial buildings, while the West side is mainly occupied by an urban park (Parc de l'Esplanade). In 2015, d'Auteuil Street counted 71 properties, two-third of which are condominiums, and the others rental apartments, offices and hotels. The street itself was first designed in the second half of the 18th century, but was settled mostly from the mid-19th century by bourgeois citizens and wealthy traders. The choice of reorganizing the landscaping of d'Auteuil Street can be explained by its 'posh' character and its proximity to the Parliament. D'Auteuil Street had been among the very first areas to benefit from the 'new' model of preservation created in the early 1980s: between 1980 and 1999, two-thirds of the properties located on the street had been restored.



Note: The figure presents the d'Auteuil Street (*Rue d'Auteuil*) configuration. While the East side of the street is occupied by residential and commercial buildings, the west side is occupied by an urban park (*Parc de l'Esplanade*). The park was already present before the street reorganization.

Source: Google Map, 2016.

Figure 7.2: D'Auteuil street

By the summer of 2005, d'Auteuil Street was reorganized following an agreement between the CCNQ and Quebec's city council. The restoration ended in September of the same year and was conducted in two main streams. First, it comprised some measures regarding urbanism. Among them was the reorganization of the traffic into a one-way street leading south of the historic district. At the same time, the road was narrowed into a two-lane road instead of four before the reorganization occurred (see figures 7.3a and 7.3b)

Second, the landscaping of the street was underlined by a heritage policy. Some work was done to replace asphalt by a traditional paved road. Sidewalks were enlarged and decorated with plants and other vegetation. Lastly, but not least, monuments dedicated to foreign poets such as Alexander Pushkin, Dante Alighieri and Komitas were installed (Figure 7.3b). Such a heritage policy mirrored the wish to underline the international character of the city of Quebec according to the CCNQ's mission. At the same time, d'Auteuil Street remained a part of the birthplace of New France, as put forth by the main heritage policy that has prevailed since the 1960s. This case study shows that heritage in a historic district is by no mean cast in stone. The meanings given to the historic district can change through the years, as heritage policies themselves change. Hence, heritage policies are dynamic variables hedonic modeling has to turn to. The costs and profitability of such policies should be understood in order to properly implement them since they are intended not only to transform urban landscape but mainly to underpin historic sites' symbols and heritage.

Figure 3a. The new D'auteuil Street



Source: CCNQ

Figure 3b. Monuments dedicated to foreign poets



Source: CCNQ

Figure 7.3: D'Auteuil Street after restoration

7.2.4 Heritage Policies and Hedonic Modeling of Residential Property Values in Historic Districts

Generally speaking, urban greenspace and tree cover projects seem to have a positive impact on property values (see for example, Roy et al., 2012, Sander et al., 2010). However, the D'auteuil Street reorganization is not limited to greenspace. As a part of a patrimonialization project, the associated heritage policy having a cultural and symbolic dimension is desired. As noted by Ryberg-Webster and Kinahan (2014: 123) there is “a need for more empirical studies on the contemporary relationship between preservation and urban revitalization and planning.”

Several scholars have already attempted to measure the impacts of a historic district designation on residential property values through hedonic modeling. Most of the time, the investigations are limited to the premium that follows the localization of a residential property in a historic district. The historic designation is expected to have positive impacts on property values by increasing neighborhood prestige and neighborhood social capital or by providing a protection from construction projects (Rickman, 2009).

Most authors have pointed out a positive impact associated with a historic designation. Narwold et al. (2008) studied the case of San Diego (CA) and found a 16% premium on house prices, while Clark and Herrin (1997) and Leichenko et al. (2001) also noted higher property values, respectively for Californian and Texan case studies. The same result is found by Asabere and Huffman (1997) in Philadelphia (PA) and by Coulson and Leichenko (2001) in Abilene (TX).

However, others have also shown that historic designation can decrease house prices or have a modest effect (Coffin, 1989). This may be explained by the pressures being put on owners by laws and regulations related to the historic districts (Rickman, 2009). Heintzelman and Altieri (2013) found that if benefits may be associated with the creation of

a historic district, they are overwhelmed by the negative effects of the associated restrictions. This is also supported by Asabere et al. (1994) who considered the case of Newburyport (MA) and found that property values are influenced by the architectural style.

In between are those who have pointed out the distinctive impact of historic designation on residential property values depending on the types of residential properties in terms of house prices. For example, Zahirovic and Chatterjee (2012) found that the positive impact of historic preservation is mainly observed for the properties sold at the lower end. However, Asabere et al. (1994) suggest the inverse by showing that the smaller historic apartment buildings experienced a reduction of the selling price.

These previous results outline different contextual factors and the complexity of historic districts. Most of the time, upon investigating residential property values in historic districts, heritage policy effects, other than the designation act itself, are ignored. Of course, designation is an important part of heritage policies, though by no means could it be considered as the only heritage policy at work in a historic district (Berthold, 2012a). Only Noonan (2007) and Leichenko et al. (2001) have introduced this pattern, measuring the impacts of a private property historic designation on hedonic modeling. In all, the conclusions suggest at least a positive, and in most cases significant, effect on house prices. However, the way the historic dimension is expressed in the independent variables is not well developed: it neglects the possibility that the effect may well depend on punctual urban landscaping policies.

Even if cultural heritage has been described as being intangible, and therefore not clearly measurable, the use of hedonic modeling allows assessing and estimating an economic value using implicit prices related to extrinsic amenities. Beyond the fact of being located inside a historic district, proximity to historical sites of interest, carrying a symbolic dimension, motivates the willingness to pay for such location. Moro et al. (2013) observed that distance to cultural heritage sites positively influences housing prices in

Dublin, Ireland. For their part, van Duijn and Rouwendal (2013) observed that proximity to cultural heritage sites is particularly valued by wealthier households. Lazrak et al. (2014) attempted to measure such intangible heritage factors by assessing the effect of listed buildings on market price. Their results show a 26.9% increase for a listed building's price and a 0.28% average increase for the surrounding buildings in a 50-meter radius. However, for their part, Ahfeldt and Meannig's conclusion (2010: 314) suggests that "the totality of the built environment [...] constitutes the amenity recognized by real estate markets."

Overall, in the old city context heritage policies are meant to sustain the symbolic construct of the area. According to literature, proximity to a heritage site is positively valued by the residential market as well as the location in a designated area. Despite the contribution of punctual heritage policies, to our knowledge, no previous work attempted to assess how housing values react to an urban development project in the context of a historic district. Our work proposes to investigate the effect of a particular heritage policy through modification in landscaping *inside* the historic designated district. According to previous studies' conclusions, if the considered heritage policies successfully managed to become the place of commemoration and the cultural symbol sought by the CCNQ,⁴² positive impacts on property values are expected.

7.2.5 Methodological Framework

Hedonic theory (Rosen, 1974) suggests that the price of a complex good can be expressed as a function of the various attributes forming the bundle. Hedonic modeling is appropriate for cross-sectional analyses (Gibbons and Machin, 2008) and is extensively

⁴² <http://www.capitale.gouv.qc.ca/realisations/voies-d-acces/rue-d-auteuil> (accessed in, August, 2016).

applied in real estate literature. In a multiple linear regression framework, this means that each coefficient related to a given characteristic represents its implicit (hedonic) price. In the case of housing, many amenities, both intrinsic and extrinsic to the property, are proven to play a significant influence on the determination of the market value of units.

In such a case, the sale price of a real estate good j sold at time period t is represented by a vector p_{jt} of dimension $[N_T \times 1]$ where N_T corresponds to the sample size. In order to control for heteroskedasticity issues and to ensure that the dependent variable is normally distributed, it is common practice to express the sale price using a logarithmic transformation (Dubé et al., 2011a). The sale price is usually expressed as being a function of a set of intrinsic and extrinsic amenities. These amenities are considered through matrices \mathbf{X}_{jt} and \mathbf{Z}_{jt} , respectively of sizes $[N_T \times K]$ and $[N_T \times M]$, where K corresponds to the number of intrinsic characteristics considered and M is the number of extrinsic amenities. Since real estate prices are subject to be influenced by temporal market fluctuations a matrix \mathbf{D}_{it} of dimension $[N_T \times (T - 1)]$ composed of temporal dummy variables accounts for nominal prices variations (Equation 7.1).

$$p_{jt} = \alpha + \mathbf{D}_{jt}\boldsymbol{\delta} + \mathbf{X}_{jt}\boldsymbol{\beta} + \mathbf{Z}_{jt}\boldsymbol{\theta} + u_{jt} \quad (7.1)$$

Where $\boldsymbol{\delta}$, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ are vectors of parameters respectively of dimension $[(T - 1) \times 1]$, $[K \times 1]$ and $[M \times 1]$, while the parameter α refers to the constant term of the equation.

Since the main objective of the paper is to evaluate the impact of a given change in the environmental amenities on house prices after a given time period, the equation 7.1 is augmented by including two new components measuring the possible effects related to proximity of the given place, as well as the change in this effect after a given date.

First, a set of variables controlling for the distance between a sold good j and the place where the change occurs, R_{jc} is developed (Equation 7.2). Four different distance buffers ($C= 1, \dots, 4$), based on Euclidean buffers, are considered within 0 to 600 meters (Equation 7.2), with the over 600 meters as being the reference.

$$\begin{aligned}
 R_{j1} &= \begin{cases} 1 & \text{if good } j \text{ is within a } 0 - 150 \text{ m radius} \\ 0 & \text{otherwise} \end{cases} \\
 R_{j2} &= \begin{cases} 1 & \text{if good } j \text{ is within a } 150 - 300 \text{ m radius} \\ 0 & \text{otherwise} \end{cases} \\
 R_{j3} &= \begin{cases} 1 & \text{if good } j \text{ is within a } 300 - 450 \text{ m radius} \\ 0 & \text{otherwise} \end{cases} \\
 R_{j4} &= \begin{cases} 1 & \text{if good } j \text{ is within a } 450 - 600 \text{ m radius} \\ 0 & \text{otherwise} \end{cases}
 \end{aligned} \tag{7.2}$$

Another indicator variable is introduced to identify the moment when a change occurs, at time t^* . For instance, this change is related to the implementation of the heritage policy related to D'Auteuil Street, which corresponds to the end of the construction in September 2005 (Equation 7.3).

$$\tau_{jt} = \begin{cases} 1 & \text{if good } j \text{ is sold at } t \geq t^* \\ 0 & \text{otherwise} \end{cases} \tag{7.3}$$

A Hadamard product, a term-by-term multiplication, is used to isolate the effect of proximity depending on the period of the sale (before/after end of the street construction).

The new variables are included in the original hedonic pricing model (Equation 7.1) to obtain the difference-in-differences (DID) hedonic model (Equation 7.5).

$$\mathbf{p}_{jt} = \boldsymbol{\iota}\alpha + \mathbf{D}_{jt}\boldsymbol{\delta} + \mathbf{X}_{jt}\boldsymbol{\beta} + \mathbf{Z}_{jt}\boldsymbol{\theta} + [\boldsymbol{\tau}_{jt}\odot\mathbf{R}_{jc}]\boldsymbol{\gamma}_c + \mathbf{u}_{jt} \quad (7.5)$$

Where, $\boldsymbol{\tau}_{jt}$ is a vector of dimension $[N_T \times 1]$, taking the value of 1 if the transaction occurred after the end of the construction (September, 2005) and 0 otherwise. The dimension of the matrix \mathbf{R}_{jz} is $[N_T \times C]$, corresponding to the number of considered buffers and $\boldsymbol{\gamma}_c$ is the corresponding vector of parameters of dimension $[C \times 1]$ associated with the post-renovation effect for each of the z radiuses.

The model with the DID estimator can be estimated using this functional form. However, there are still many challenges related to the use of hedonic pricing models (Dubé and Legros, 2014c). Thus, the DID estimator is transformed using a repeated sales (RS) approach for two main reasons. First, using a first difference of the price equation allows to adequately isolate the impact of a change in intrinsic or extrinsic amenities, while controlling for omitted variables that influence the price determination process. The DID estimator in a repeated sales approach has the advantage of cancelling the effect of constant environmental elements, such as other spatial amenities (shops, other historic sites, etc.) or condominium characteristics⁴³, considering that they remain stable over time, unlike the considered street restoration. Assuming that the d'Auteuil Street renovation is part of an urban development policy decided and implemented independently of the economic agents (residential real estate buyers and sellers), the modification can be considered exogenous, at least from the perspective of both agents (Dubé et al., 2011b, Gibbons and Machin, 2008).

⁴³ However, the model controls the effect of obtaining subsidies for renovations.

And second, real estate transactions are known to be spatially autocorrelated, which violates one of the assumptions of the ordinary least squares (OLS) approach and results in imprecise estimators and variance. The spatial autocorrelation problem may arise due to model misspecification and omitted variables issues. Being free from functional form (McMillen, 2010) and omitted variables problems (Dubé and Legros, 2014b), a DID estimator based on repeated sales is a proper way to overcome such limitations (Dubé et al., 2011b).

The repeated sales approach is based on transactions of a given good that appears at least twice over the time period under consideration. This transformation necessarily reduces the total sample size available for estimating the problem, reducing the size from N_T to n_T , with $n_T \ll N_T$. Denoting by s the moment of the first transaction (sold) and by r the moment of the second transaction (resale) and taking the first difference between price equation at each moment returns the final equation to be estimated using the repeated sales approach (equation 7.6).

$$\begin{aligned}
 (\mathbf{p}_{jtr} - \mathbf{p}_{jts}) &= (\boldsymbol{\alpha}_r - \boldsymbol{\alpha}_s) + (\mathbf{D}_{jtr} - \mathbf{D}_{jts})\boldsymbol{\delta} + (\mathbf{X}_{jtr} - \mathbf{X}_{jts})\boldsymbol{\beta} \\
 &\quad + (\mathbf{Z}_{jts} - \mathbf{Z}_{jtr})\boldsymbol{\theta} + ([\boldsymbol{\tau}_{jtr} \odot \mathbf{R}_{jz}^*] - [\boldsymbol{\tau}_{jts} \odot \mathbf{R}_{jz}^*])\boldsymbol{\gamma}_z \\
 &\quad + (\mathbf{u}_{jtr} - \mathbf{u}_{jts})
 \end{aligned} \tag{7.6}$$

Which can be simplified since the new independent variable expresses the difference within the amenities regarding the moment of the transaction, assuming that the implicit price is constant over time (equation 7). The last equation can be simplified as shown in Equation 7.7 and where ε_j corresponds to the new *i.i.d.* perturbation term, a vector of dimension $[N_T \times 1]$.

$$\Delta \mathbf{p}_j = \Delta \mathbf{D}_j \boldsymbol{\delta} + \Delta \mathbf{X}_j \boldsymbol{\beta} + \Delta \mathbf{Z}_j \boldsymbol{\theta} + [\Delta \boldsymbol{\tau}_j \odot \mathbf{R}_j^*] \boldsymbol{\gamma}_z + \varepsilon_j \tag{7.7}$$

Where the vector of dependent variable, $\Delta \mathbf{p}_j$, is now of dimension $[N_T \times 1]$, while the matrix of time dummies variables indicating the moment of sale ($\mathbf{D}_{js} = -1$) and resale ($\mathbf{D}_{jr} = 1$), $\Delta \mathbf{D}_j$, is of dimension $[N_T \times (T - 1)]$, while the two other matrices of independent variables expressing the difference in the intrinsic ($\Delta \mathbf{X}_j$) and extrinsic ($\Delta \mathbf{Z}_j$) amenities are, respectively, of dimension $[N_T \times K^*]$ and $[N_T \times M^*]$, where K^* is the total number of intrinsic amenities that vary over time, while M^* is the total number of extrinsic amenities that vary over time. The matrix $[\Delta \boldsymbol{\tau}_j \odot \mathbf{R}_j^*]$ allows to isolate the impact of the change related to the historical policy related to d'Auteuil Street, and is of dimension $[N_T \times C]$. As before, the vectors of parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ permit the retrieval of the implicit prices of varying amenities, the vector $\boldsymbol{\delta}$ allows to capture for the nominal aspect of price, while the vector $\boldsymbol{\gamma}_z$ allows to isolate the impact of the heritage policy on sale prices.

The repeated sales approach is based on three assumptions. First, by using such an estimator one postulates that the good's sales frequency follows a random process, meaning that no particular factor can explain why a good is sold more often than another (Gatzlaff and Haurin, 1997, 1998)⁴⁴. Second, if no particular information is available, the amenities of the real estate goods are assumed to be constant, otherwise the change in the amenities is captured by a vector of variables ΔX_{kjrS} (Dubé et al., 2011b). Finally, the specification assumes that the coefficients are constant over time (Case, 1997, Case and Shiller, 1989a).

⁴⁴ Gatzlaff and Haurin (1997:48) outline that in practice: "Because only a small percentage of houses typically sell each year (or some other measured interval), the sample of houses that sell may have nonrandom statistical properties which can bias the index estimate".

7.2.6 Empirical application and transaction data

The dataset has been kindly transmitted by the real estate valuation service of the City of Quebec and collected using the land register of the Province of Quebec.³⁴⁵ It originally consists of 1,631 transactions of condominiums occurring between 1997 and 2011. A few transactions have been discarded to ensure that: i) the sale price is available; ii) the exact location of the transactions is available. After discarding some transactions due to missing or imprecise information, the dataset consists of a total of 1,477 transactions from which 744 repeated sales are identified and compose the final sample. The mean sale price is \$132 666 CAD while the mean resale price is \$178 170 CAD, which corresponds to a 34% increase of the mean nominal prices (Table 7.1).

Table 7.1: Temporal distribution of the transactions

	Sale price			Resale price		
	<i>Transactions</i>	<i>Mean</i>	<i>std. dev.</i>	<i>Transactions</i>	<i>Mean</i>	<i>std. dev.</i>
1997	34	113,893.6	75,708.6	1	89,466.7	37,302.19
1998	98	93,645.9	51,698.1	25	101,119.0	68,257.2
1999	77	99,782.1	72,701.7	22	96,140.4	64,781.8
2000	54	94,260.7	45,620.8	27	95,881.5	37,339.0
2001	90	108,993.5	74,126.1	45	91,278.7	35,206.3
2002	65	116,412.8	63,598.4	48	116,019.2	76,254.2
2003	81	153,670.2	72,315.7	56	147,042.3	83,906.2
2004	82	163,358.6	70,419.7	74	153,439.2	71,049.4
2005	57	171,928.7	74,616.1	84	177,085.6	73,810.1
2006	33	162,166.7	72,012.3	52	186,351.3	88,012.3
2007	27	194,854.9	97,539.7	58	216,689.6	103,985.3
2008	22	204,240.5	60,760.8	70	206,222.8	75,429.5
2009	21	182,160.7	48,694.2	71	237,587.6	99,944.7
2010	2	390,000.0	21,213.2	82	255,417.5	93,022.5

⁴⁵ After estimations, the transactions price indexes progression corresponds to the ones observed for the entire Québec City region.

	Sale price			Resale price		
	Transactions	Mean	std. dev.	Transactions	Mean	std. dev.
2011	1	151,000.0	-	27	237,461.2	87,128.2
Overall	744	132,666.8	769,66.91	744	178,170.8	96,251.24

Using the geographical coordinates of each transaction, a total of 97 different buildings are identified (same pairs of coordinates). The transactions occur in some buildings spatially concentrated within the walls of the Old City (Figure 7.1), while the small number of independent buildings where transactions occur suggests that many transactions are located in the same building. In fact, the data reveals that a maximum of 71 transactions have been noted in the same exact location (i.e., in the same building).

A particularity needs to be noted for the 'inside the walls' part of the Old City. First, due to its age, the old city district's structure of Québec City is hardly comparable to modern urban structure. While Euclidean distance may be appropriate to represent a form of proximity in most of modern North American district configurations, the present case study is composed of dense and tortuous streets that could affect the notion of proximity. Moreover, historic districts are usually places where walking is valued. For this reason and for comparative purposes, the estimation of the model is based on both Euclidean and network distances.⁴⁶ Second, the location of the street that experiences a change related to the historical policy is located along an urban park and close to the walls. Thus, there are no transactions that occur on the West side of the street, i.e., there is no building on this side (Figure 7.2).

⁴⁶ The Euclidean distance (d_{ij}) is computed using: $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. Where x and y represent the coordinates of the observations i and j . The street is considered as a line and the Euclidean distance can be defined as the shortest path from a particular property to this line. Regarding the network distance, GIS software (ArcMap) is used to compute the shortest road distance between a property and the street.

Table 7.2 presents the spatial distribution of the transactions for the different radiuses used within walking distance from d’Auteuil Street. On the one hand, 184 transactions are included within a 600-meter radius when a network measure of distance is used. The number of transactions occurring before the project was 110, while 74 transactions occurred after. On the other hand, 337 transactions are located in the same radius when the Euclidean measure of distance is chosen, which suggests that both distances express different proximities. Of this total, 203 transactions were recorded before the street renovation and 134 after. This first comparison shows how the network distance captures a more “local” effect by including fewer transactions within the same buffers.

Table 7.2: Spatial distribution of the transactions: proximity to the renovated street

Network distance	# of transactions			Euclidean distance	# of transactions		
	<i>All</i>	<i>Before</i>	<i>After</i>		<i>All</i>	<i>Before</i>	<i>After</i>
<i>0-150m</i>	43	22	21	<i>0-150m</i>	47	26	21
<i>150-300m</i>	55	38	17	<i>150-300m</i>	89	58	31
<i>300-450m</i>	49	30	19	<i>300-450m</i>	74	47	27
<i>450-600m</i>	37	20	17	<i>450-600m</i>	127	72	55
<i>> 600m</i>	560	368	192	<i>> 600m</i>	407	275	132
Total	744	478	266	Total	744	478	266

7.2.7 Main findings and estimation results

Considering that the DID estimator is produced using a first difference mathematical transformation, the model is estimated without a constant term, which alters the usual interpretation of the R^2 statistic (Dubé et al., 2013b, Wooldridge, 2012). For this reason, the

performance of the model is assessed using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

The comparison (base) model (model 1, Table 7.3) is estimated using only the quarterly dummy variables determining the trend of nominal price evolution between 1997 and 2011. Here, the only information being taken into account is related to changes in price appreciation related to the renovations of building. The results show that the prices more than doubled within a decade, between 2003 and 2011. The same trend has been observed for the Québec urban agglomeration (Dubé et al., 2011b). However, price appreciation is not statistically linked to the fact that owners have received a grant for undertaking some renovations between the time of the sale and the resale.

The second and third models are extensions of the base model adding spatial variables capturing the effect of the historical policy work depending on the proximity to d'Auteuil Street and the moment of the transaction (sale before and resale after street renovations). At this stage, it is important that there is no need to include information on all other intrinsic and extrinsic amenities if these amenities do not change over time. This largely reduces the number of independent variables in the models, while adequately controlling for these aspects using the first difference transformation of the DID estimator hedonic price equation. In the second model, network distance is used to measure the proximity to the policy-targeted portion of the street, while the third model use network distances. The new specification presents lowered AIC and BIC statistics, suggesting that the additions of such variables contribute to increase the quality of the model.

According to the coefficients associated with the spatial variables, the price variation for properties located within 0 and 150 meters from d'Auteuil Street (0 to 150m, network distance) is not influenced by the landscaping policy. However, a significant effect is observed for the two following radiuses (150 to 300m and 300 to 450m), while the effect is not significant over 450 meters. For the two significant radiuses, the results suggest a price

depreciation varying between \$13,432 CAD and \$10,812 CAD, respectively, as compared to the more distant properties.

The third model is estimated using a similar specification to Model 2, but with Euclidean distance-based buffers. According to the lowered BIC and AIC statistics, the model quality is again increased as compared to Model 1. As observed in the previous specification, the properties within 150 meters of the street do not experience a significant price appreciation related to the renovation of the street. However, for properties located within 150 and 300 meters, a negative and significant effect is again observed. In contrast to the network distance variables, the significant effect is limited to the second buffer when Euclidean distance is used and corresponds to an \$11,268 CAD price decrease⁴⁷, a comparable impact to the previous specification.

Overall, the results point to three main conclusions. First, the assistance programs for renovations of properties located in Old Quebec have no impact on the selling price variations. This does not mean that condominium renovations are not valued. It is conceivable that renovated condominiums are sold at higher prices, meaning that public funds used for private renovations would be transformed in a greater profit margin for the seller and to a lesser part in taxes revenues for the municipality. However, this is not the case here. One assumption would be that such renovations remove some “cachet” from the original apartments, even when offering some additional and more contemporaneous commodities. Another assumption would be that the effect is minimized considering that most of the condominiums in the area have been well restored a few years before the analysis. In fact, as previously mentioned, two-thirds of the condominiums have been restored by the owner between 1980 and 1999.

⁴⁷ The price depreciation is obtained using the product of the estimated significant coefficient and the mean price variation (\$45,504 CAD).

Second, the results suggest a localized and negative effect of the landscaping renovation of d'Auteuil Street for condominiums located within 150 to 450 meters network distance and within 150 to 300 meters Euclidian distance. As a matter of fact, the reorganization and landscaping of d'Auteuil Street has brought more road traffic to this part of the district than previously, but more over noise pollution according to the change in the material used to build the road. The reorganization of the street into a one-way road leading south also added high traffic volume in the heart of the historic district (more than 400 vehicles every hour from 7–9 am and 3–5 pm on work days) and several studies have already shown how noise can lead to a depreciation of residential values, as such might be the case with airports and housing values (Cohen and Coughlin, 2008, Baranzini and Ramirez, 2005). In the current situation, noise as a negative externality could be a result of the pavement of the road itself (made of stones and not of asphalt) on which vehicles travel up to 50km/h (31m/h).⁴⁸ Unlike asphalted roads, a paved road generates an uneven friction between the car wheels and the ground, generating more noise.

Another negative externality could be related to the redesigned parking spaces location. Surrounding streets may experience the drawback of the new street design since parking spaces are pretty rare for residents inside the walled section of the Old City. In fact, the new one-way street may also change parking habits for vehicles heading north. This assumption seems relatively plausible considering the scarcity of parking spaces in such dense and tortuous areas.

Yet, even if the closest properties should be the more influenced by the renovation work, no significant effect is noticed on their resale price. Thus, for properties located right along the way, or very close, the positive externalities, attributed to heritage factors and

⁴⁸ Sandberg (1987) has investigated the issue of road traffic noise in regard to different surface components, namely stones and asphalt.

visual aspect, the landscaping work and the addition of vegetation as well as construction of larger sidewalks can counterbalance the negative externalities. In fact, one can assume that the new vegetation could only have a minor impact since most of it is added on the west side of the street, where an urban park was already present. The condominiums therefore certainly do not benefit from much of the plant cover (noise reduction, intimacy, shadow). Interestingly, a minimized impact of vegetation would support the presence of a positive impact of intangible heritage factors. However, the expanded sidewalk and the new monuments are located on the west side of the street (opposite side of the buildings). The new noisy and busy street organization is then close to the residential properties.

Third, the results outline the stability of the results by using two different definitions of the proximity. According to the descriptive statistics, the use of network distance presents a more concentrated distribution of the observations between distance radiuses, and a more “local” definition of proximity. In fact, most of the observations identified in a 300 to 450m network distance buffer are located in the 150 to 300 meter Euclidean distance buffer. The fact that both results return similar conclusions gives an interesting robustness check. Both models suggest a price depreciation varying between \$13,432 CAD and \$10,812 CAD.⁴⁹

⁴⁹ Additional robustness tests have been performed using continuous distances and applying an inverse distance function as well as a negative exponential function. In both cases, the coefficient associated with the distance is not significant. These results can be explained by the non-linear decay functions applying larger weights to closest observations not affected by the change. This test using continuous distance also confirms the choice of buffers for a more precise measure in this particular case.

Table 7.3: Estimation results

<i>Independent variables</i>	Model (1)		Model (2)		Model (3)	
	<i>Coef.</i>	<i>Sig.</i>	<i>Coef.</i>	<i>Sig.</i>	<i>Coef.</i>	<i>Sig.</i>
1997-Q2	<i>Reference</i>		<i>Reference</i>		<i>Reference</i>	
1997-Q3	-0.2288		-0.2385		-0.2299	
1997-Q4	0.5699	***	0.5566	***	0.5608	***
1998-Q1	-0.5747	**	-0.5879	**	-0.5826	**
1998-Q2	-0.0553		-0.0752		-0.0713	
1998-Q3	-0.0183		-0.0479		-0.0453	
1998-Q4	0.0503		0.0254		0.0340	
1999-Q1	0.0686		0.0196		0.0192	
1999-Q2	0.1000		0.0627		0.0690	
1999-Q3	0.0303		0.0210		0.0208	
1999-Q4	0.1623		0.1242		0.1351	
2000-Q1	0.1597		0.1397		0.1463	
2000-Q2	0.0783		0.0675		0.0660	
2000-Q3	0.1142		0.1130		0.1129	
2000-Q4	0.1507		0.1292		0.1434	
2001-Q1	0.1242		0.0999		0.1142	
2001-Q2	0.1082		0.0884		0.0909	
2001-Q3	0.1863		0.1611		0.1826	
2001-Q4	0.1813		0.1626		0.1709	
2002-Q1	0.2496		0.2309		0.2354	
2002-Q2	0.3019	*	0.2969	*	0.3000	*
2002-Q3	0.2132		0.1955		0.2072	
2002-Q4	0.3549	**	0.2879	*	0.3142	*
2003-Q1	0.2500		0.2431		0.2505	
2003-Q2	0.5753	***	0.5585	***	0.5674	***
2003-Q3	0.3297	*	0.2783	*	0.2904	*
2003-Q4	0.6495	***	0.6340	***	0.6331	***
2004-Q1	0.4444	***	0.4320	***	0.4328	***
2004-Q2	0.5782	***	0.5693	***	0.5823	***
2004-Q3	0.5748	***	0.5434	***	0.5683	***
2004-Q4	0.6266	***	0.6094	***	0.6181	***
2005-Q1	0.6875	***	0.6722	***	0.6814	***

<i>Independent variables</i>	Model (1)		Model (2)		Model (3)	
	<i>Coef.</i>	<i>Sig.</i>	<i>Coef.</i>	<i>Sig.</i>	<i>Coef.</i>	<i>Sig.</i>
2005-Q2	0.6989	***	0.6855	***	0.6922	***
2005-Q3	0.7318	***	0.7406	***	0.7152	***
2005-Q4	0.6736	***	0.6771	***	0.6647	***
2006-Q1	0.6978	***	0.7290	***	0.7047	***
2006-Q2	0.6741	***	0.6681	***	0.6591	***
2006-Q3	0.7554	***	0.7540	***	0.7331	***
2006-Q4	0.7701	***	0.7803	***	0.7594	***
2007-Q1	0.8387	***	0.8398	***	0.8070	***
2007-Q2	0.8567	***	0.8792	***	0.8712	***
2007-Q3	0.8403	***	0.8573	***	0.8271	***
2007-Q4	0.9031	***	0.9078	***	0.9005	***
2008-Q1	0.8385	***	0.8540	***	0.8377	***
2008-Q2	0.9492	***	0.9573	***	0.9276	***
2008-Q3	1.0158	***	1.0422	***	1.0145	***
2008-Q4	0.9313	***	0.9123	***	0.8963	***
2009-Q1	0.7928	***	0.7868	***	0.7669	***
2009-Q2	0.8729	***	0.8946	***	0.8788	***
2009-Q3	0.8376	***	0.8351	***	0.8054	***
2009-Q4	1.0667	***	1.0831	***	1.0760	***
2010-Q1	0.9712	***	0.9812	***	0.9726	***
2010-Q2	1.1202	***	1.1241	***	1.1060	***
2010-Q3	0.9001	***	0.9329	***	0.9170	***
2010-Q4	1.1378	***	1.1314	***	1.1238	***
2011-Q1	1.1329	***	1.1350	***	1.1015	***
2011-Q2	1.0868	***	1.0993	***	1.0848	***
2011-Q3	1.3397	***	1.5400	***	1.5564	***
Grant for renovations (Binary)	0.0386		0.0231		0.0202	
<i>Impact - distance to d'Auteuil Street</i>						
Network dist., 0 to 150 m (Buffer)			-0.0405			
Network dist., 150 to 300 m (Buffer)			-0.2952	**		
Network dist., 300 to 450 m (Buffer)			-0.2376	*		
Network dist., 450 to 600 m (Buffer)			0.1353			
Network dist., over 600 m			<i>Reference</i>			

<i>Independent variables</i>	Model (1)		Model (2)		Model (3)	
	<i>Coef.</i>	<i>Sig.</i>	<i>Coef.</i>	<i>Sig.</i>	<i>Coef.</i>	<i>Sig.</i>
Euclidean dist., 0 to 150 m (Buffer)					0.0102	
Euclidean dist., 150 to 300 m (Buffer)					-0.2476	**
Euclidean dist., 300 to 450 m (Buffer)					0.0883	
Euclidean dist., 450 to 600 m (Buffer)					0.0795	
Euclidean dist., over 600 m					<i>Reference</i>	
<i>N</i>	744		744		744	
<i>R2</i>	0.575		0.586		0.585	
<i>Akaike Information Criteria (AIC)</i>	692.977		682.129		683.867	
<i>Baysian Information Criteria (BIC)</i>	960.475		968.075		969.813	

Legend: * p<0,05; ** p<0,01; *** p<0,001

7.2.8 Conclusion

Designated by law as a protection area and inscribed on the World Heritage List, UNESCO, the historic district of Old Quebec is an important realm of memory not only for the people of Québec but also for North America, since it was the birthplace of New France. As part of a patrimonialization approach, heritage policies help to maintain and support a cultural and ideological heritage that may be transposed to residential values through location factors and proximity effects. Previous studies have concluded that both heritage or cultural factors and location in an old city district are valued by the residential market. Thus, the heritage dimension of patrimonialization approach could be expected to be an intangible source of positive externalities. However, while heritage policies are implemented in an old city district with non-concealed expectations regarding symbolism building, no previous work attempted to assess their impacts on property values.

The paper proposed to assess the impact of a particular heritage policy implemented in an old city context: the reorganization of d'Auteuil Street, located in the designated old Quebec City district, with a heritage policy transforming the landscape of the street in 2005. The impact is assessed using property values (condominiums sold between 1997-2011) and hedonic price modeling. A repeated sales approach based on a difference-in-difference

(DID) estimator is applied to isolate the effect of proximity to the considered street while controlling for the impact of constant extrinsic and intrinsic property-related factors.

While positive effects could be expected due to heritage and landscaping factors, the results underline a complex pattern. Based on two different definitions of proximity to the street, network distance and Euclidean distance, the results underline a different conclusion, with no significant impact detected for direct proximity (along the street) and negative impact for residential buildings located close to the street. Estimation results also outline that both distance definitions show similar and therefore robust results. Yet, the network distance appears to be a more precise and appropriate definition of proximity in such a context.

The fact that no significant difference is noticed before and after the renovations for immediate proximity properties (0 to 150m) doesn't mean that the proximity to d'Auteuil Street is not valued. It only means that renovations do not, generally speaking, bring about a positive net plus-value. We can therefore reasonably assume that the improvement generates a situation where the positive externalities are completely outweighed by the negative externalities, at least in this particular real estate market. In other words, from the residential market perspective the positive expected impacts of the heritage policy in terms of patrimonialization and symbolic construction could be cancelled by physical landscaping decisions. The negative impact on greater distance properties may be explained by a set of negative externality drawbacks associated with landscaping and urban reorganization decisions.

Of course, the global economic impact of the renovation of the street may clearly depend on other factors, such as the tourism effect, but it is not accounted for here. The impact could also be different for other real estate types, such as commercial buildings and offices. Thus, the calculation of the total economic impact of such heritage policies should not be simplified to only the condominium market. However, the results show that creating

places that attract more people and that generate more noise may also induce some cost for the local resident, as measured by the price decrease on the real estate market. Overall, such negative effect needs to be relativized, but clearly underlines the potential collateral negative effect related to heritage landscape policies.

Finally, even if the analysis is limited to one particular case study, the results suggest that heritage conservation policies are complex in nature due to the equilibrium between intangible and physical landscaping dimensions. In fact, this suggests that any heritage policy implemented in a historic district context cannot be taken separately from the general urban context and development strategies. In such a situation, policy makers and planners should be aware that the expected economic impacts could vary from the traditional perspective. In the actual case, the observed decreasing property values suggest that landscaping strategies are not profitable in the long term, at least from the residential property taxes viewpoint. It should be mentioned that expected economic impacts of such heritage policies may not only affect tourism and that possible interrelations between economic residential impacts and tourism impacts remain to be explored.

CHAPITRE 8
EFFET D'ANTICIPATION ET POST-CONSTRUCTION DE L'EXTENSION
D'UNE LIGNE DE MÉTRO SUR LES VALEURS RÉSIDENIELLES : LE CAS
DE LAVAL (CANADA), 1995-2013

8.1 PRÉSENTATION DE L'ARTICLE

Ce quatrième et dernier article est une seconde étude de cas qui intègre la notion de dépendance spatiale dans un contexte spatiotemporel. L'article, intitulé « Anticipation and post-construction impact of a metro extension on residential values : The case of Laval (Canada) 1995-2013 », a été réalisé en collaboration avec les professeurs Jean Dubé et Philippe Apparicio. La recherche a fait l'objet d'une présentation au congrès annuel de la *Mid-continental regional science association* (MCRSA) de 2015 à Saint-Louis, MI (É.U.).

Son objectif est d'évaluer l'impact de nouvelles installations de transport en commun (extension du métro de Montréal vers Laval) sur les valeurs résidentielles, tout en intégrant les notions d'effets de débordements spatiaux dans un contexte temporel. Ce travail se positionne de manière complémentaire à l'article précédent. La méthodologie proposée repose également sur une approche par ventes répétées et sur l'estimation d'un modèle DID. Néanmoins, alors que le modèle DID permet d'intégrer certains effets spatiaux, il ne permet pas d'intégrer les relations de dépendance associées aux prix de vente. Une variante spatiale du modèle est donc proposée, SDID (Spatial Difference-in-Differences), afin d'intégrer cette composante spatiale supplémentaire (Dubé et al., 2014).

L'évaluation des impacts de ces infrastructures n'est pas chose aisée (Handy, 2005, Higgins and Kanaroglou, 2016a). De manière générale, les travaux scientifiques se concentrent sur l'évaluation de l'impact des infrastructures de transport en fonction de la distance (Debrezion et al., 2007). Pourtant, l'effet pourrait ne pas être homogène le long de la ligne de transport et fluctuer selon les stations de métro (Hess and Almeida, 2007). Certains suggèrent également que l'effet n'est pas constant dans le temps et que des effets d'anticipation ou spéculatifs peuvent être présents avant même la mise en place du nouveau service de transport (Agostini and Palmucci, 2008, Atkinson-Palombo, 2010, McDonald and Osuji, 1995, McMillen and McDonald, 2004). L'article propose donc d'évaluer et de décomposer l'effet de l'extension d'une ligne de métro par période (annonce du projet, construction, mise en service) ainsi que par station.

Les résultats de l'étude suggèrent que l'effet du nouveau service de transport en commun n'est pas spatialement homogène. Alors qu'une analyse globale (sans distinction des stations) permet d'identifier un effet positif associé à la proximité du métro après sa mise en service, une analyse décomposant l'effet par station indique que l'effet est associé à une seule des trois nouvelles stations. En revanche, cet effet n'est pas non plus homogène dans le temps. L'effet s'avère significatif après l'ouverture au public du service de transport, mais un effet d'anticipation est également observé durant la période de construction.

La mise en place de l'expansion du réseau souterrain de métro dans la région de Montréal est une action onéreuse. L'évaluation de l'impact de manière globale et sans distinction des différentes périodes d'implémentation ne permettrait pas d'en évaluer justement les retombées. De tels projets motivent généralement de grandes attentes en termes de retombées. Les résultats de la recherche invitent à en relativiser les effets et à envisager le développement de réseaux de transport moins coûteux que par le rail.

8.2 ANTICIPATION AND POST-CONSTRUCTION IMPACT OF THE METRO EXTENSION ON RESIDENTIAL VALUES: THE CASE OF LAVAL (CANADA), 1995-2013

8.2.1 Abstract

The application of hedonic pricing models has a long history in estimating a premium associated with infrastructure related to public transportation. However, results accuracy largely relates to methodological and empirical considerations such as: i) presence of spatial latent component (spatial autocorrelation); ii) temporal breaks related to different periods over which the infrastructure is built; and iii) heterogeneity of the effect along the line and stations. The paper aims to assess the impact of Montréal's metro extension to the city of Laval (announced in 1998 and started operating in 2007). A spatial difference-in-difference (SDID) estimator based on a repeated sales approach is used to isolate the impact of the proximity to the new infrastructure on single-family house prices. The results show a non-homogenous impact of the infrastructure depending on the implementation phases and the stations. Moreover, an anticipation effect is observed for this same station but is not apparent when the analysis uses a global perspective (all stations combined).

8.2.2 Introduction

Land values are a good indicator of mobility (Burgess, 2008). In the Alonso-Muth-Mills model (Alonso, 1964, Muth, 1969, Mills, 1969), it is explicit that a reduction in transportation costs can be positively translated into land and property values. In consequence, location rent can be isolated using proximity to mass transit lines or stations, depending on the type of public transportation. Yet, it must be admitted that the assessment of the expected impacts by planners and policymakers remains a complex task (Higgins and Kanaroglou, 2016b). Despite abundant literature on the topic, varying results are observed between studies (Debrezion et al., 2007).

From a methodological perspective, real estate values are location dependent (Can, 1992, Dubin, 1998), suggesting that latent spatial components are likely to result in spatial autocorrelation among residuals of pricing equations (Lancaster, 1966, Rosen, 1974), invalidating standard regression assumptions of independence (Legendre, 1993, Anselin and Griffith, 1988) and, by extension, conclusions of empirical applications. Spatial autocorrelation is a central concept in empirical spatial research (Getis, 2008) and is defined as the coincidence between measures depending on location (LeSage and Pace, 2009a, Anselin and Bera, 1998). It may be associated with the omission of significant spatial variables (McMillen, 2010) or with spatial dependence processes structuring a data generating process (Le Gallo, 2002).

Other methodological challenges emerge when trying to isolate the effect of accessibility to a mass transit (MT) system on real estate values. The impact may be heterogenous and varies over time according to the implementation phases of the development of new MT lines, but also among the stations serving the line. The construction of transportation infrastructures can be broken down into distinct implementation phases (announcement, construction, operation) during which the impacts

on real estate values can fluctuate according to anticipation effects of the market. Stated otherwise, the impact measured during the planning periods can well be different from the final impact according to possible speculation or anticipation effects (Agostini and Palmucci, 2008, Atkinson-Palombo, 2010, McMillen and McDonald, 2004, McDonald and Osuji, 1995). While a global assessment of the effect is often realised for the whole transportation system, station effects are also likely to generate localised premium distinctions (Hess and Almeida, 2007) due to station characteristics, services quality or landscaping and urban development choices.

The paper aims to assess the impact of the extension of Montreal's subway on the (suburban) island of Laval (on the north side) using 5,422 pairs of transactions collected between 1995 and 2013. A repeated-sales approach (Case and Shiller, 1987, Bailey et al., 1963) is used and matched with a spatial difference-in-difference model (SDID) (Dubé et al., 2014) to deal with possible spatial omitted variables and control for spatial spillovers over transaction prices. For comparison purposes, two model specifications are estimated. The first assesses the premium associated with each project implementation phase from a global perspective (all stations combined) while the second breaks down this periodic effect for each new station. Additional statistical tests for homogeneity and cumulative significance effect over the implementation phases are performed. The results show evidence of a proximity effect when all stations are combined after the first phase (announcement). Moreover, the impact is not homogenous along the transport line, with only positive significant effect being estimated for one of the three stations.

The paper is divided into five sections. The first section is dedicated to literature review presenting previous empirical studies and paying special attention to the variation of study results depending on the project implementation phases and contextual factors. The second section covers the proposed modelling approach. The third section presents the data

used to estimate the model, while the fourth section presents the estimation results. The last section covers a discussion and a conclusion closes the paper.

8.2.3 Literature review

The bid-rent theory (Alonso, 1964, Muth, 1969, Mills, 1969) suggests that land prices are a decreasing function of distance to the center (or central business district - CBD) (O'Sullivan, 2007). According to the complex structure of actual cities, theory now needs to be considered according to realities such as urban sprawl and polycentric forms. In this context, the development of the transportation infrastructures contributes to supplant the notion of physical proximity to the benefit of accessibility and modifies the distribution of socio-economic profiles across urban space (Glaeser et al., 2008). According to Burgess (2008), an increased mobility acts as a catalyst of change in a city and land values are “one of the most sensitive indexes of mobility” (idem: 344) explaining the consideration for hedonic pricing modelling using property values. High expectations are associated with mass transit systems in terms of urban and economic developments (Dittmar and Ohland, 2004; Hess and Lombardi, 2004; Landis et al., 1994; Calthorpe, 1993). Such infrastructures' implementation process takes a considerable amount of time and money to be built. However, since the early work of Dewees (1976) and Bajic (1983), our capacity to predict those development policies' impacts is still limited (Handy, 2005) and the extensive literature reflects great results variability due to temporal and spatial considerations.

Agostini and Palmucci (2008) suggested that the variability is associated with new development of MT systems in three distinct phases: announcement period; construction period; and operation period. The announcement period refers to the moment the project is publicized. During this period a speculation effect may be observed indicating a first market response (anticipation) to an MT system development. The second phase, the

construction period, marks the realization of the project. At that time the location of the future stations is concretely observed by the public, where the anticipation effect can be exacerbated. McMillen and McDonald (2004) considered the Chicago transit system and pointed out that house prices were affected even before the stations were built but after the project plans were known. Bae et al. (2003) also identified an anticipatory effects for the construction of a subway line in Seoul. Knaap et al. (2001) and Atkinson-Palombo (2010) identified capitalisation benefits from the time the project was publicly known. Studying Santiago's metro system, Agostini and Palmucci (2008) have noted that the average apartment price rises after the project has been announced, and that there is smaller, but positive and significant, impact after the identification of the station location. On the opposite side, Yan et al. (2012) did not observe any anticipation effect of the project proximity on home prices before the rail system began operation in Charlotte (North Carolina).

After the MT comes into operation, the maturity of the service is another plausible source of impact variation. While a number of studies note a positive impact of fully developed light rail systems on real estate values (Weinstein et al., 2002, Cervero and Duncan, 2002), Gatzlaff and Smith (1993) did not find any relationship for a semi-developed rail system. Mohammad et al. (2013) illustrate those conclusions by citing the example of the *Metropolitan Atlanta Rapid Transit rail system* (MARTA) whose impact was first studied 10 years after its initial operation by Nelson and McCleskey (1989). The authors only found a minimal effect while, a few years later Bowes and Ihlanfeldt (2001) identified positive and negative effects depending on the location. On the opposite side, Cervero and Landis (1997) assessed the impact of the *Bay Area Rapid Transit* (BART) system during its first 20 years of operation. The results for the fully developed service are consistent with those previously obtained after the first operation (Dyett et al., 1979).

Of course, spatial dimension is also important. Each station is different and can produce a distinctive impact (Hess and Almeida, 2007). In fact, from an urban development perspective, each station can be perceived as a singular infrastructure with service, structural and landscaping particularities. The premium measured may not be homogenous over the entire transportation system's line and vary due to station effects. Landis et al. (1994) analysed five different heavy and light rail transit systems in six California counties and outlined that the service capitalization affects home prices for several reasons, mainly including service quality (reliability, speed, frequency). Some authors even found that the frequency of the service is more important than the proximity to the transit service (Debrezion et al., 2011). According to others, the complementarity between different transportation modes also affects nearby values (Voith, 1993, So et al., 1997, Ryan, 2005). As exposed by Bowes and Ihlanfeldt (2001) the presence of parking facilities positively affects the sales prices. By studying Hamburg's transit system, Brandt and Maennig (2012a) state that subterranean stations significantly increase the impact on condominium prices. The location of the station has also been noted as a factor influencing the premium. According to Cervero (2006) and Mulley et al. (2016), the most important premium is for residential and commercial values near downtown stations, with the impact varying among the MT system, with the higher effect related to commuter rail services (Debrezion et al., 2007). Of course, the effect may not be linear over space (Chen et al., 1997), but follow an inverse U-shape. Empirical works noted that effect of proximity may be lowered by negative externality predominance such as noise, pollution or criminality (Diaz and Mclean, 1999, Bowes and Ihlanfeldt, 2001, Bajic, 1983). Ihlanfeldt (2003) outlined the complexity of such effects by finding positive and negative impacts of Atlanta's MARTA transit system on crime depending on location. Hess and Almeida (2007) find similar results for the city of Buffalo. Other transportation modes may even impact the premium due to substitution effects (Baum-Snow et al., 2005, Ryan, 1999).

Considering temporal and spatial consideration, the decomposition and isolation of a proximity effect of a new transportation station per project implementation phase and station implies devising an appropriate econometric model.

8.2.4 Methodology and research strategy

Hedonic theory and price function is one of the most used methodologies to break down the value of a complex good into its characteristics' implicit prices (Rosen, 1974). The approach is based on revealed preferences: actual sale price reflects an equilibrium reached where both agents (buyer and seller) agree on a given amount (pay and receive) for a given bundle of goods according to the composition of the bundle. Widely documented in the urban and real estate literature, the hedonic pricing model therefore allows modelling the price of a real estate good as a function of its intrinsic and extrinsic attributes. The former relate to the elements describing the property sold while the latter are associated with environmental or location factors.

The log-linear functional form is described as one of the most used and best linear specifications (Dubé et al., 2011a, Dubé et al., 2013b). The hedonic pricing model expresses the sale price (in logarithmic transformation) of a property i at time t , stacked in a vector \mathbf{y}_{it} of dimension $[N_T \times 1]$, where N_T corresponds to the total number of observations ($N_T = \sum_{t=1}^T N_t$), as a function of the amenities of the goods, intrinsic and extrinsic, stacked in matrices \mathbf{X}_{it} and \mathbf{Z}_{it} , respectively of sizes $[N_T \times K]$ and $[N_T \times M]$, where K is the total number of intrinsic amenities and M is the total number of extrinsic amenities. To account for the fact that prices are usually expressed in nominal terms, a set of temporal dummy variables, stacked in a matrix \mathbf{D}_{it} of dimension $[N_T \times (T - 1)]$, is usually added to the model specification. Finally, it is current to explicitly account for a

spatial spillover effect in real estate markets by using a spatial autoregressive specification based on the spatial lag of the dependent variable (Equation 8.1).

$$\mathbf{y}_{it} = \alpha + \mathbf{W}\mathbf{y}_{jt}\rho + \mathbf{D}_{it}\boldsymbol{\delta} + \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{Z}_{it}\boldsymbol{\theta} + \boldsymbol{\xi}_{it} \quad (8.1)$$

Where \mathbf{W} is a row-standardized spatial weights matrix, of dimension $[N_T \times N_T]$, and the resulting spatial variable, $\mathbf{W}\mathbf{y}_{jt}$, expresses the mean sale price of properties in the direct vicinity. The vectors of parameters, $\boldsymbol{\delta}$, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ are, respectively, of dimension $[(T - 1) \times 1]$, $[K \times 1]$ and $[M \times 1]$, with $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ measuring the implicit prices of the amenities, while the parameters α and ρ are scalars measuring, respectively, the constant term and the strength of the transaction price's spatial dependence (LeSage and Pace, 2009a).

The analysis of the impact of an “exogenous” change in extrinsic amenities can be explored using a difference-in-differences (DID) specification by identifying the moment where change occurs over time. In the case of a new MT system, the change can be broken down into four (4) distinct phases. Keeping the pre-announcement period as reference, each T_p^* period is defined by: announcement t_2^* ; construction t_3^* ; and the operation t_4^* (Equation 8.2).

$$\begin{aligned} T_2^* &= \begin{cases} 1 & \text{if } t_1^* \leq t \leq t_2^* \\ 0 & \text{otherwise} \end{cases} \\ T_3^* &= \begin{cases} 1 & \text{if } t_2^* \leq t \leq t_3^* \\ 0 & \text{otherwise} \end{cases} \\ T_4^* &= \begin{cases} 1 & \text{if } t \geq t_4^* \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (8.2)$$

Multiplying these indicator variables by the appropriate extrinsic amenities returns the spatial DID (SDID) estimator.⁵⁰

To simplify the presentation, let's stack these dummy variables in a matrix, \mathbf{T}_{it} , that identifies the moment of the sale during the p phases of the project. A Hadamard product, \odot , is used as a term-by-term multiplication isolating the effect of the change on the extrinsic amenities' matrix, where the three ($p = 3$) vectors of parameters, $\boldsymbol{\gamma}_p$, are all of dimension $[M^* \times 1]$, where M^* is the total number of extrinsic amenities changing over time and assume an homogenous effect within all stations considered. Multiplying the matrix of temporal indicators to the extrinsic amenities that change over time returns the DID estimator (Equation 8.3)

$$\mathbf{y}_{it} = \boldsymbol{\alpha} + \mathbf{W}y_{jt}\rho + \mathbf{D}_{it}\boldsymbol{\delta} + \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{Z}_{it}\boldsymbol{\theta} + [\mathbf{T}_{it}\odot\mathbf{Z}_{it}^*]\boldsymbol{\gamma}_p + \boldsymbol{\varepsilon}_{it} \quad (8.3)$$

Where \mathbf{Z}_{it}^* is the matrix of extrinsic amenities that change over time, of dimension $[N_T \times M^*]$.

Of course, and as previously mentioned in the literature review, the premium measured for proximity to a new transportation station may differ among stations. In order to assess a possible distinctive effect for each new station, a new variable (vector), \mathbf{L}_{it} , is included in the specification to isolate the closest station s from the new MT line. The matrix \mathbf{L}_{it} is composed by S different vector of dimension $[N_T \times 1]$, where a given vector, \mathbf{L}_{sit} ,

⁵⁰ Considering that spatial autocorrelation rises by dependency process between sale prices, a spatial difference-in-difference (SDID) (Dubé et al., 2014) is estimated controlling for spatial spillovers structuring independent variables.

$$\mathbf{y}_{it} = \boldsymbol{\iota}\alpha + \mathbf{W}\mathbf{y}_{jt}\rho + \mathbf{D}_{it}\boldsymbol{\delta} + \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{Z}_{it}\boldsymbol{\theta} + [(\mathbf{T}_{it}\odot\mathbf{L}_{it})\odot\mathbf{Z}_{it}]\boldsymbol{\gamma}_{ps} + \boldsymbol{\varepsilon}_{it} \quad (8.4)$$

Where the $[p \times s]$ vector of parameter $\boldsymbol{\gamma}_{ps}$ are all of dimension $[M^* \times 1]$.

The SDID estimator can be estimated using a price equation. However, it can also be estimated using a reduced form based on the repeated-sales (RS) approach. The RS approach is a well-known method to develop a price index of the evolution of real estate values (Bailey et al., 1963). It also allows to adequately control for latent spatial structure constant over time (Dubé et al., 2011b, Dubé et al., 2013b). Moreover, the SDID estimator using the RS approach has the advantage of being exempt from problems related to the functional form (McMillen, 2010) and omitted variables (Dubé et al., 2014).

However, the DID and SDID estimator relies on three assumptions (Dubé et al., 2014). First, it assumes that the coefficients are constant over time (Case and Shiller, 1987, Case and Shiller, 1989b), meaning that preferences remain unchanged. Second, it considers, if no information is available to describe changes in amenities, that intrinsic amenities are constant over time (Dubé et al., 2011b). Finally, the SDID estimator postulate that the frequency of the sale of a good follows a random process suggesting that no characteristic affects the sale frequency of a property (Gatzlaff and Haurin, 1998, Gatzlaff and Haurin, 1997).

Since repeated transactions are only a fraction of total transactions, sample size is reduced ($n_T \ll N_T$). Individual observations consist of houses first sold in time f and with the resale in time r . The SDID estimator is obtained by using the first difference from the hedonic pricing model according to the moment of the sale.

Depending on the specification used (equations 8.3 or 8.4), the SDID based on repeated sales approach (using a first difference equation) returns a different and reduced form (equations 5 or 6, respectively).

$$\Delta \mathbf{y}_i = \Delta \mathbf{W} \mathbf{y}_i \rho + \Delta \mathbf{D}_i \delta + [\Delta \mathbf{T}_i \odot \mathbf{Z}_i] \boldsymbol{\gamma}_p + \xi_i \quad (8.5)$$

$$\Delta \mathbf{y}_i = \Delta \mathbf{W} \mathbf{y}_i \rho + \Delta \mathbf{D}_i \delta + [\Delta \mathbf{T}_i \odot (\mathbf{L}_i \odot \mathbf{Z}_i)] \boldsymbol{\gamma}_{ps} + \xi_i \quad (8.6)$$

Since intrinsic amenities are assumed to be unchanged between the sale and resale periods, their effect is cancelled by the first difference ($\Delta \mathbf{X}_i = \mathbf{0}$). Moreover, the change in extrinsic amenities is captured through the multiplicative (DID) term, resulting also in cancelling the vector of extrinsic amenities ($\Delta \mathbf{Z}_i = \mathbf{0}$). The matrix $\Delta \mathbf{y}_i$ corresponds to the difference in log of sale price (and approximation of price growth) and is of dimension $[n_T \times 1]$, where n_T is the total number of pairs of transactions.⁵¹ Similarly, $\Delta \mathbf{W} \mathbf{y}_i$ corresponds to the variation of the autoregressive parameter (spillover effect) and is of dimension $[n_T \times n_T]$.

The elements of the matrix $\Delta \mathbf{D}_i = (\mathbf{D}_{ir} - \mathbf{D}_{if})$, indicating the month of the sale ($\mathbf{D}_{if} = 1$) and resale ($\mathbf{D}_{ir} = 1$), can thus take a value of 1, 0 or -1. By the same token, the exogenous changes in the period of the implementation process are captured by the matrix $\Delta \mathbf{T}_i = (\mathbf{T}_{ir} - \mathbf{T}_{if})$ of dimension $[n_T \times p]$. The elements of the matrix take the value of 1 during the first sale ($\mathbf{T}_{if} = 1$) as well as at the moment of the resale ($\mathbf{T}_{ir} = 1$). Thus, as for the $\Delta \mathbf{D}_i$ matrix, the elements of the matrix $\Delta \mathbf{T}_i$ can take a value of 1 (a change occurs in the resale period), a value of 0 (no change occurs between moment of sale and resale), or a

⁵¹ Since repeated transactions (couples of sale and resale) are only a fraction of the total transactions, we should have $n_T \ll N_T$.

value of -1 (a change occurs during the sale period). The general elements of the matrix $(\mathbf{L}_i \odot \mathbf{Z}_i)$, of dimension $[n_T \times s]$, take a value returning the distance to the closest station for transactions occurring at least twice. Finally, $\boldsymbol{\varepsilon}_{it}$ is a vector of error terms of dimension $[n_T \times 1]$ and assumed independent and identically distributed (*iid*).

8.2.5 Case study and data

The Laval metro extension project

The city of Laval includes 369 000 inhabitants, located in the northern suburbs of Montréal and is therefore part of the greater metropolitan area in the province of Québec. For the urban agglomeration of Montréal, over 51% of the population is living in Laval but nearly 53% work on Montréal Island mainly in the service sector (82%).⁵² The general configuration of the whole city region is distinguished by the presence of the river physically isolating Laval and Montreal islands from the North and South shores (Figure 8.1).

⁵² Reference: *Statistics Canada*; census of 2011.



Figure 8.1: City map and spatial distribution of the transactions

Before the extension of Montreal’s subway (orange line) to Laval, bridges and suburban trains allowed access to Montreal Island. According to the policy makers, the project of the extension of the actual subway line aimed to decongest roads, increase residents' quality of life, while being part of a “sustainable development” approach from an environmental perspective. The new transportation stations were first announced by the *Parti Québécois*, a provincial political party in power at the time, in October 1998, for an approximate cost of \$179M. Originally the project was designed for a single station but two additional stations were added. In 2001 corrections were made to the original project so that the planned costs were subsequently re-evaluated at \$379M. The total length of the orange

line is now about thirty kilometers and has 31 stations, with the new section being slightly over 5 km in length (Figure 8.2).

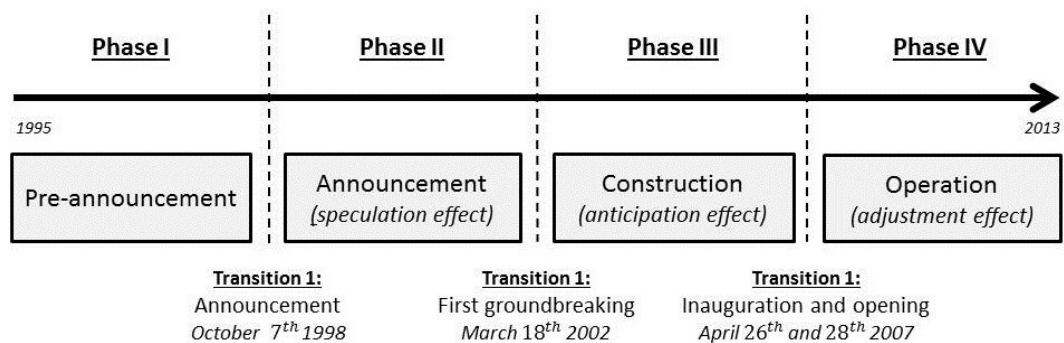


Figure 8.2: Project implementation timeline

The first ground was broken in March of 2002. After the election of the *Parti Libéral du Québec*, another provincial political party, in 2003 and with the beginnings of the construction, the cost continued to rise, to reach a total of \$803.6M. From the total cost, \$785.3M was reimbursed by the governments (provincial and federal) and \$18.3M by the *Agence Métropolitaine de Transport* (AMT). However, after the completion and first operation of the subway in April 2007, some confusion remains regarding the total project costs.

The closest station to Montréal, *Cartier station*, is an intermodal facility with a connection to the bus system and 465 parking spaces. A part of the station is dedicated to small shops. The second station, *De la Concorde station*, is also an intermodal facility with connection to the suburban train. Landscaping around the stations was done, to include gardens. Finally, the *Montmorency station* also features a link to bus lines and 1 357 indoor and outdoor parking spaces, but it is distinguished from other stations being the terminus of

the metro line. Since the end of the work, new expansion projects are regularly planned but not yet underway.

Property transaction data

The transactions data are obtained from the *Greater Montréal Real Estate Board (GMREB)* and originally consist of 22,160 residential property (N_T) transactions between 1995 and 2013 in the city of Laval. Each observation contains information on the transaction price and addresses allowing to geocode each property at its exact coordinates using a geographic information system (GIS). From the total number of transactions, the observations with incomplete data or selling price information are rejected. The final sample consists of 5,422 pairs of transactions (sale and resale - n_T) used for the repeated sale hedonic modelling. The distribution of sales and resales per period is presented in Table 8.1.

Table 8.1: Yearly distribution of sale and resale transactions (1995 to 2013 – 5612 pairs)

Resale	Sale																		Total	
	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012		
1996	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	Phase I
1997	8	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	
1998	11	17	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	
1999	15	33	29	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	93	
2000	23	54	41	30	30	3	0	0	0	0	0	0	0	0	0	0	0	0	181	
2001	28	54	52	38	51	24	2	0	0	0	0	0	0	0	0	0	0	0	249	
2002	13	43	46	62	61	67	58	3	0	0	0	0	0	0	0	0	0	0	353	Phase III
2003	9	37	43	57	71	74	91	48	6	0	0	0	0	0	0	0	0	0	436	
2004	10	37	45	40	56	47	63	85	61	8	0	0	0	0	0	0	0	0	452	
2005	8	29	31	47	60	44	37	68	92	42	3	0	0	0	0	0	0	0	461	
2006	12	28	39	40	44	40	32	60	66	55	24	8	0	0	0	0	0	0	448	
2007	2	7	5	11	14	4	9	11	17	20	12	0	0	0	0	0	0	0	112	
2008	6	22	17	28	33	25	23	44	54	50	32	30	6	8	0	0	0	0	378	Phase IV
2009	6	20	33	20	29	26	24	53	55	54	44	51	7	25	9	0	0	0	456	
2010	12	18	21	32	41	21	22	39	61	62	38	49	16	44	37	11	0	0	524	
2011	8	17	23	24	35	19	27	31	52	44	29	46	9	39	44	32	4	0	483	
2012	7	25	22	24	29	18	24	33	46	34	30	23	17	41	71	43	21	10	518	
2013	0	2	0	2	2	4	2	4	9	3	2	2	3	6	2	3	1	2	49	
Total	178	451	453	471	557	416	414	479	519	372	214	209	58	163	163	89	26	12	5,244	
	Phase I				Phase II				Phase III				Phase IV							

Note: Each shade of grey represents a particular phase of the implementation period

Table 8.2: Temporal distribution of transaction

	Sale price			Resale price		
	Transactions	Mean	std. dev.	Transactions	Mean	std. dev.
1995	178	95,963.4	27,838.1	0	0	0
1996	451	98,110.4	28,849.7	2	27,250	3,181.98
1997	453	98,932	28,075.6	14	94,164.3	14,110
1998	471	99,913.8	29,727	35	103,929	37,689.6
1999	557	102,431	30,894.9	93	97,378.5	24,049.5
2000	416	105,979	30,531.6	181	104,498	28,766.4
2001	414	110,596	30,798.4	249	111,190	29,678.2
2002	479	132,230	37,628.4	353	133,270	39,574.4
2003	519	159,627	39,136.8	436	153,819	34,617
2004	372	187,711	53,712	452	177,122	39,733.4
2005	214	188,220	41,218.1	461	193,113	46,765
2006	209	212,216	57,786.4	448	206,812	51,689.3
2007	58	226,346	61,280.9	112	223,913	57,793.9
2008	163	236,628	61,259.7	378	235,824	56,469.7
2009	163	244,973	65,688.4	456	246,976	66,393.8
2010	89	253,512	71,671.5	524	263,242	65,449.5
2011	26	273,989	76,110.5	483	278,024	66,973.5
2012	12	203,392	23,847.7	518	290,476	69,879.6
2013	-	-	-	49	302,782	65,562.8
Total	5,244	138,164	62,765.1	5,244	209,931	80,428

The mean sale price is \$138,164 (CAD) while the mean resale price is \$209,931. According to the distribution of the sales and resales for each year (Table 8.2), the total number of sales is relatively limited during the first years of the dataset corresponding to the pre-announcement period. This number increases slightly during the announcement period and grows more markedly for the subsequent periods (construction and operation phases). A total of 656 sales and 2,814 resales experienced the new transportation service by being observed after April 2007 (operation phase). Similarly, respectively 1,985 and 2,058 sales and resales are observed during the construction period while 1,583 sales and 360 resales are recorded between the announcement and the beginning of the construction work (Table 8.2).

By using the exact location of the properties, the *Euclidean* distance (d_{ij}) to the closest metro station is computed, and the inverse distance ($1/d_{ij}$) is used as an independent variable.⁵³ Considering the first transactions (sales), the mean distance is higher for properties sold during the announcement phase (11,341 m.), while the lowest distance is observed during the construction phase (119 m.). Considering the second transactions (resales), the same mean distance, maximum and minimum mean distances, are respectively observed during the construction phase and after the first operation of the service (Table 8.3).

⁵³ The Euclidean distance (d_{ij}) is computed using: $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. Where x and y represent the coordinates of the observations i and j .

Table 8.3: Distance to the nearest metro station per project phase

	Sales				Resales			
	N	Mean	Min	Max	N	Mean	Min	Max
Cartier station								
<i>Pre-announcement</i>	180	3996,87	168,832	7253,751	2	2272,10	1820,68	2723,521
<i>Announcement</i>	248	4280,45	251,836	7125,292	57	4199,47	168,832	6989,469
<i>Construction</i>	308	4213,77	206,603	7157,793	335	4180,66	227,13	7125,292
<i>Operation</i>	103	3696,45	217,907	7006,838	445	4078,94	175,438	7253,751
De la Concorde station								
<i>Pre-announcement</i>	811	6294,23	488,545	11109,98	8	6377,34	4080,21	8179
<i>Announcement</i>	1283	6288,61	402,534	11341,94	291	6475,42	762,197	11109,98
<i>Construction</i>	1631	6482,10	471,851	11305,7	1669	6334,30	471,851	11341,94
<i>Operation</i>	528	6069,66	611,285	11083,2	2285	6320,65	402,534	11228,45
Montmorency station								
<i>Pre-announcement</i>	29	1141,29	279,475	2045,088	2	846,91	515,272	1178,544
<i>Announcement</i>	52	1226,42	211,366	2397,061	12	1184,43	279,475	1883,637
<i>Construction</i>	46	1073,97	119,414	1977,739	54	1188,99	272,702	2246,312
<i>Operation</i>	25	1333,20	716,953	2397,061	84	1184,42	119,414	2397,061
All stations								
<i>Pre-announcement</i>	1020	5742,31	168,832	11109,98	12	4771,40	515,272	8179
<i>Announcement</i>	1583	5807,72	211,366	11341,94	360	5938,70	168,832	11109,98
<i>Construction</i>	1985	6004,81	119,414	11305,7	2058	5848,72	227,013	11341,94
<i>Operation</i>	656	5516,53	217,907	11083,2	2814	5812,83	119,414	11228,45

8.2.6 Results

A global perspective

As previously mentioned, two models are estimated: i) one that assumes a homogenous effect within stations among the different periods (Equation 8.5); and ii) one that allows different effect for each stations among the different periods (Equation 8.6). The first specification allows the temporal decomposition of the proximity impact to the new stations according to the period of the implementation process, while the second allows for a temporal decomposition of the proximity impact on house prices as well as for spatial decomposition according to the station. Both models are estimated using a spatial two-stage generalized least squares (GS2SLS)⁵⁴ controlling for spillover effect between properties which could lead to bias in the estimated coefficients' variance.⁵⁵

The SDID transformation simplifies the model specification but has the particularity of eliminating the constant term, affecting the interpretation of the R^2 statistic (Wooldridge, 2012). The model's performance is then assessed using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

Keeping the first quarter of 1995 as the reference period, the price index fluctuations of the sale/resale properties over the considered period show that market trends experience

⁵⁴ Details on the GS2SLS estimations can be found in Drukker et al. (2013).

⁵⁵ As described in the methodology section, the spatial weight matrix is restricted to the transactions occurring during the same period and is specified using an inverse distance function. According to LeSage and Pace (2014), the choice of the distance function has no impact on the modelling estimations. Different weights matrix functions (negative exponential of binary) were tested. The results remain robust for the different specifications.

a 102% to 107% increase of the price between sale and resale depending on the model specification (Figure 8.3).

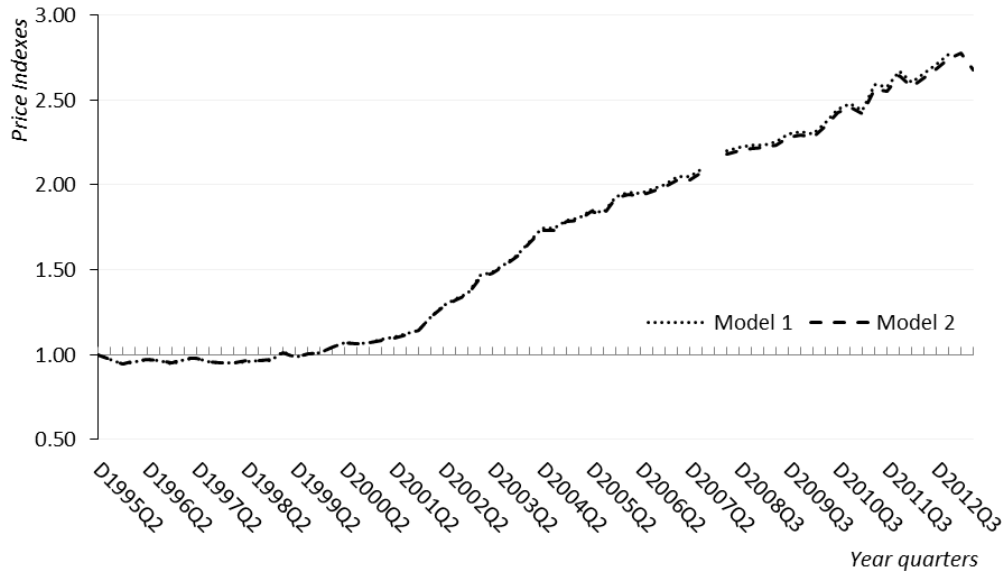


Figure 8.3: Sale/Resale price evolution (1995-2013)

Assuming a homogenous effect for all the stations (model 1), proximity to the new stations significantly and positively influences the (log of) sale price ratio (0.0344, Table 8.4 and Figure 8.4)⁵⁶ only after it comes into operation (the pre-announcement period acts as reference). Unlike results suggested in some other case studies (Gibbons and Machin,

⁵⁶ Here, the decision is made only regarding the coefficient associated with the proximity of the stations regarding the different phases. For a full interpretation of the marginal effect, the Average Total Impact (ATI), referring to LeSage and Pace (2009: 36-37), should be computed.

2008, McMillen and McDonald, 2004, Bae et al., 2003) no anticipation effect is measured: the only impact comes from the beginning of the operation after the opening of the stations.

The effect is concentrated in the last period (operation phase), but a statistical test performed to evaluate the global impact ($\gamma_{T_1} + \gamma_{T_2} + \gamma_{T_3} = 0$) suggests no significant impact (Table 6a). Stated otherwise, even if an effect is measured in the last period, the previous impact phases cancelled out the impact related to the transportation system's first operation. Thus, this issue raises the question of its uncertain durability over time and could challenge some expectations in terms of economic benefits in the long term.

Table 8.4: Estimation results

Independent variables	SDID†		SDID†	
	Model (1)		Model (2)	
	<i>Coef.</i>	<i>Sig.</i>	<i>Coef.</i>	<i>Sig.</i>
Temporal dummy variable	YES		YES	
Inverse distance to the nearest station				
<i>Announcement period</i>	-0.0078			
<i>Construction period</i>	0.0133			
<i>Operation period</i>	0.0344	*		
Inverse distance to the Cartier station				
<i>Announcement period</i>			0.0138	
<i>Construction period</i>			-0.0175	
<i>Operation period</i>			-0.0180	
Inverse distance to the De la Concorde station				
<i>Announcement period</i>			0.0003	
<i>Construction period</i>			0.0500	*
<i>Operation period</i>			0.0839	**
Inverse distance to the Montmorency station				
<i>Announcement period</i>			-0.0203	
<i>Construction period</i>			0.0106	
<i>Operation period</i>			0.0136	
ρ	0.0438	***	0.0446	***
N	5244		5244	
Neighbours				
<i>Mean</i>	19.5		19.5	
<i>Min</i>	1		1	
<i>Max</i>	56		56	

Note: Heteroscedasticity has been corrected in all models;

† Inverse distance weights matrix with a mean distance threshold;

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 8.5: Statistical tests on coefficients:

Table 8.5a – Test of a global effect (models 1 and 2)

Periods	GLS (Model 1) [†]		GS2SLS (Model 2) [†]	
	Pr > chi2	sig.	Pr > chi2	sig.
All period aggregated	0.4293		0.3692	

[†] H0: no significant effect obtained when aggregating the coefficients.

Table 8.5b – Test of homogeneity per period and significance (models 3 and 4)

Periods	GSL (model 3)		GS2SLS (model 4)	
	Homogeneity [‡]	Significance [◊]	Homogeneity [‡]	Significance [◊]
	Pr > chi2	sig.	Pr > chi2	sig.
Announcement period	0.4671	0.4623	0.4699	0.6215
Construction period	0.2457	0.3783	0.1422	0.1875
Operation period	0.1483	0.0956	0.0211 *	0.0067 **

[‡] H0: no distinction between the stations.

[◊] H0: no station presents a significant distinction.

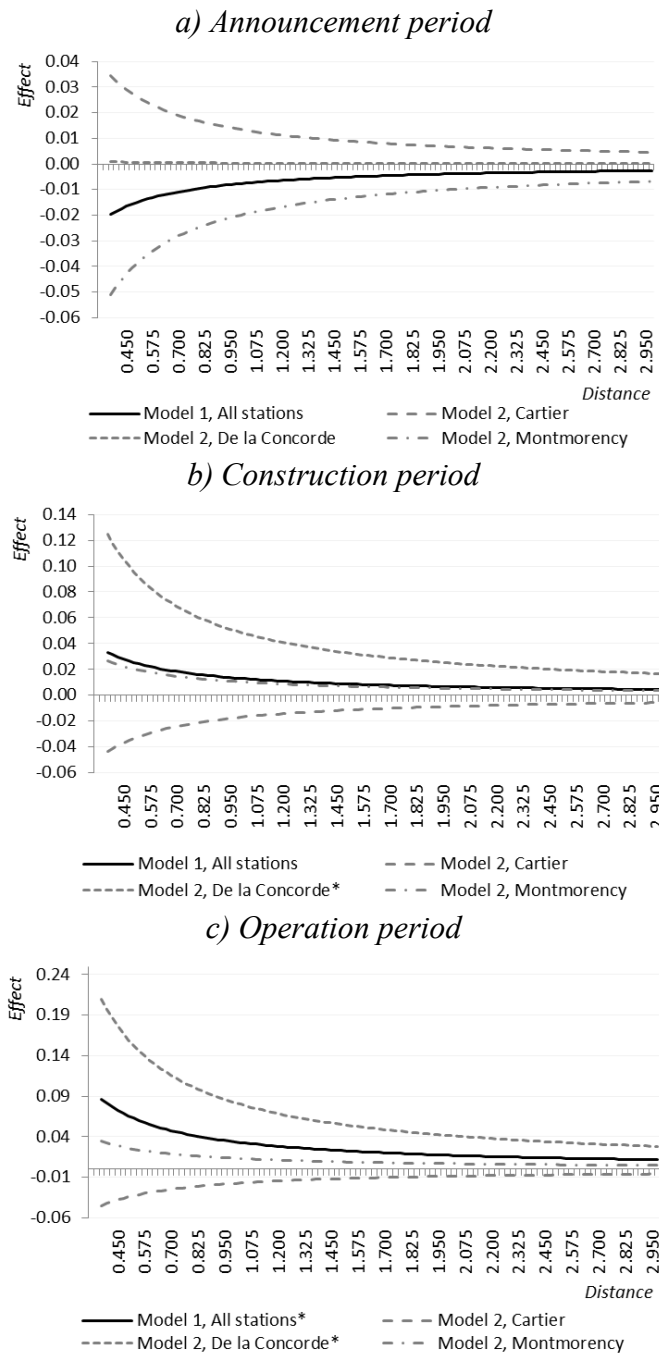


Figure 8.4: Proximity effect for each significant implementation phase

A station-oriented perspective

However, the effect can depict a more complex spatial pattern. According to Model 2, significant heterogeneity between the stations is detected: only *De la Concorde* station reveals significant impact on house price changes, with a coefficient of 0.05000 for the construction period and another one of 0.0839 during the operation period (Table 8.4 and Figure 8.4). The other stations show no significant impact on property values. According to these results, not only is the effect of proximity heterogeneous among the different implementations phases, but it also presents differences among stations. Moreover, aggregating the effect in a single and homogenous coefficient hides the impact measured for the *De la Concorde* station.

Moreover, the assumption of homogenous impact can formally be addressed using a formal statistical test within stations and periods ($\gamma_{T_1s_1} = \gamma_{T_1s_2} = \gamma_{T_1s_3}$). The results show no significant distinction within the station during the announcement and construction periods (Table 8.5b). However, the test performed on the third phase reveals a significant difference for the impact among the stations (Table 8.6b), underlining a non-homogenous effect of proximity on house prices.

Another series of tests is also computed to assess whether the global impact (sum of the three periods) per station is significant ($\gamma_{T_t s_1} = \gamma_{T_t s_2} = \gamma_{T_t s_3} = 0$ – Table 8.5b). None of the announcement periods ($\gamma_{T_1s_1} = \gamma_{T_1s_2} = \gamma_{T_1s_3} = 0$) nor construction period ($\gamma_{T_2s_1} = \gamma_{T_2s_2} = \gamma_{T_2s_3} = 0$) presents significant results. However, evidence of a significant effect is noted for the last phase ($\gamma_{T_3s_1} = \gamma_{T_3s_2} = \gamma_{T_3s_3} \neq 0$), which leads to a different conclusion than assuming a homogenous effect.

Calculating the impact on house prices

The spatial specification used (SDID) allows capturing and measuring the effect of spatial spillovers. The spillover effect needs to be accounted for when calculating the marginal effect since it acts as a chain reaction influencing prices of other houses sold at the same time period. Thus, the interpretation of the regression parameters can no longer be interpreted as the direct marginal effect on property prices unless the autoregressive parameter, ρ , is non-significant (Dubé et al., 2014).

With the standard OLS model, the marginal effect computation is straightforward and refers to the proportion of change in the dependent variable for one unit change in the independent variable (Equation 8.7). However, using a spatial specification this marginal effect also includes the induced effect of a change in all the neighbouring dependent variables, y_{it} . As presented by Dubé et al. (2014), the total marginal effect in a spatial context is then obtained by the following derivative (Equation 8.8):

$$\frac{\partial y_{it}}{\partial X_{kit}} = \beta_k \quad (8.7)$$

$$\frac{\partial y_{it}}{\partial X_{kit}} = (I - W\rho)^{-1}I\beta_k \quad (8.8)$$

In the case of the present analysis, considering the fact that the ρ coefficient is significant in both specifications, the marginal effect computation should be based on Equation 8. Moreover, since the impact varies according to the distance to the nearest station, the marginal effect needed to be decomposed to account for both effects.

For simplicity's sake, only the last specification, the non-homogenous effect within stations and periods, is discussed. Regarding the estimation results, marginal effects for different distances, varying between 400 to 1,600 meters, are reported for the *De la Concorde* station according to the significant results only (Table 8.6). The full decomposition of the direct and total marginal effect, based on the decomposition of the spatial multiplier (Steimetz, 2010) is reported in Figure 4. The decomposition of the effect is based on the two significant periods.

Table 8.6: Marginal effect of the effect of distance in dollars (\$)

		Model 2*		Model 3b**	
		<i>Operation</i>	<i>Construction</i>	<i>Operation</i>	<i>Combined</i>
400 meters	<i>Direct effect (CAD\$)</i>	6452.72	9556.36	16744.95	26301.30
	<i>Indirect effect (CAD\$)</i>	309.26	476.30	871.39	1347.69
	<i>Total effect (CAD\$)</i>	6761.98	10032.66	17616.33	27648.99
	<i>Total variation (%)</i>	9.4	14.0	24.5	38.5
800 meters	<i>Direct effect (CAD\$)</i>	3156.92	4628.90	7933.92	12562.82
	<i>Indirect effect (CAD\$)</i>	147.97	223.39	391.36	614.75
	<i>Total effect (CAD\$)</i>	3304.89	4852.29	8325.28	13177.58
	<i>Total variation (%)</i>	4.6	6.8	11.6	18.4
1,200 meters	<i>Direct effect (CAD\$)</i>	2089.48	3053.68	5196.33	8250.00
	<i>Indirect effect (CAD\$)</i>	97.21	145.79	251.74	397.53
	<i>Total effect (CAD\$)</i>	2186.69	3199.46	5448.07	8647.53
	<i>Total variation (%)</i>	3.0	4.5	7.6	12.0
1,600 meters	<i>Direct effect (CAD\$)</i>	1561.48	2278.29	3862.99	6141.28
	<i>Indirect effect (CAD\$)</i>	72.37	108.18	185.46	293.64
	<i>Total effect (CAD\$)</i>	1633.85	2386.47	4048.45	6434.92
	<i>Total variation (%)</i>	2.3	3.3	5.6	9.0

Note: * Effect for all stations.

** Effect for the *De la Concorde* Station.

For the construction period, the direct marginal effect suggests a price increase of CAD\$9,556, while the total marginal effect suggests a CAD\$10,032 price increase, corresponding to a 14% price increase. The difference between both effects represents the indirect marginal effect, measuring the impact of the spatial spillover effect. For the period where the service opens (operation), the direct marginal effect generates an increase in house prices of CAD\$16,745, while the total marginal effect is calculated at CAD\$17,616, representing \$871 resulting from spatial spillover effect. In the end, the sum of both effects returns the final effect on house prices accounting for price appreciation during the two periods. This total impact, accounting for spatial spillover effect, is calculated at CAD\$27,649, with CAD\$1,348 resulting from spatial spillover effect.

The same decomposition can be done for other distances, while the impact reduces with higher distances (Table 8.6). For example, increasing distance by 400 meters lowers the total marginal impact on house prices to CAD\$13,178 at 800 meters, to CAD\$8,648 at 1,200 meters and to CAD\$6,435 at 1,600 meters. As previously mentioned, about two-thirds of the impact results from the last period, i.e., the opening of the service. In the end, the combined effect, reported in the fourth column of Table 7, corresponds to a variation of the gain between 38.5% and 9% within 400 to 1,600 meters. Figure 8.5 illustrates the non-linear total effects. Overall, the estimated impact of the station proximity is higher when the spatial multiplier is considered and consequently influences study results.

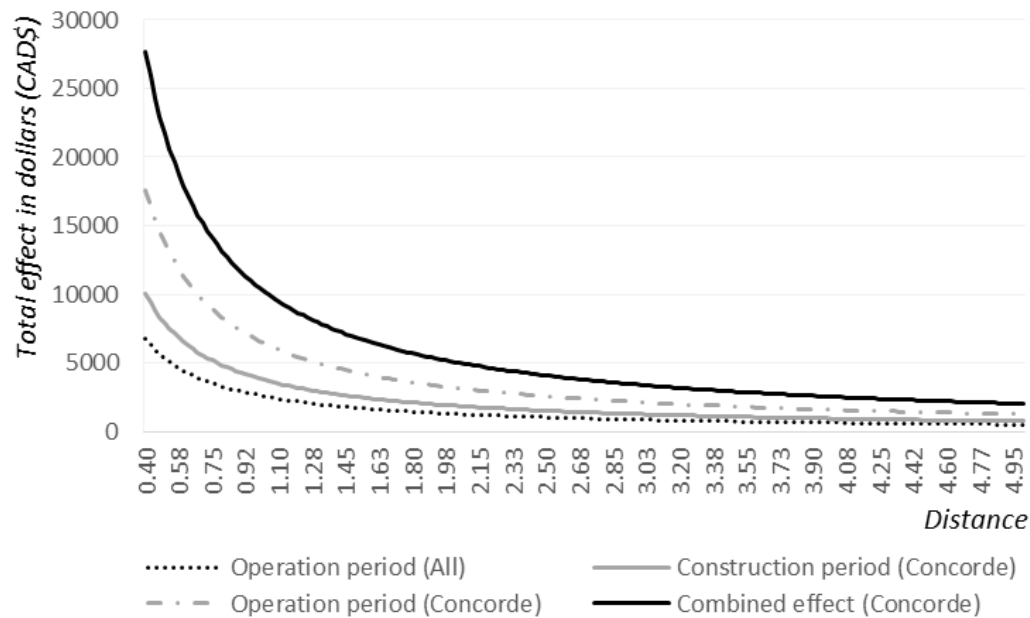


Figure 8.5: Total marginal effect expressed in dollars (CAD\$)

8.2.7 Discussion

The results show a limited effect, in space and time, for proximity to the metro stations, questioning certain expectations in terms of economics impacts, at least in the single-family market. Global measures show no significant effect after the announcement, while the decomposition of the effect per period presents a small, but positive, effect after the line comes into operation. Pushing this decomposition further, the measured impact during this period is concentrated around the *De la Concorde* station only. The results also suggest market anticipation during the construction phase for the same station.

According to these results, a global approach (all stations combined) that measures the impact reduces the significance of the results, and lowers the effect related to a given

station within the impact of other stations that shows no significant result. Thus, for the global model, assuming a homogenous impact within a station, the impact of *De la Concorde* station is hidden by the two other stations showing no significant impact. By decomposing the effect of the new infrastructure per period and stations, the estimation clearly outlined differentiated effects over space and time. A better understanding of these effects could help to better assess the economic impacts of such projects and formulate more realistic expectations according to contextual variations. Moreover, the SDID specification allows considering spatial spillovers in the total measured effect, suggesting a greater effect than suggested by coefficients.

A number of reasons could explain these variations between stations, such as contextual factors, station characteristics, but also the socio-economic status of the surrounding neighbourhoods. The premium measured is the equilibrium between positive and negative amenities evaluated by both sellers and buyers. By being the terminus of the line and located in a mainly commercial sector, with a highway and a college nearby, the coefficient associated with the *Montmorency* station may be affected by negative externalities overbalancing positive advantages of public transportation service proximity. However, it should be noted that the *Montmorency* station presents a small number of properties nearby, which can explain the absence of significant results. Even if not significant, the effect of the *Montmorency* station is negative during the announcement period and turns positive at the moment of the construction (Figure 8.4). The *Cartier* Station is located closer to a bridge connecting Laval to Montréal as compared to the other stations. A positive impact of the infrastructure could be cancelled by the density of daily commuters traveling this route, the bridge acting as a funnel. The car may also be a good alternative for residents of this area. It should be noted that, even if not significant, the impact of the station is positive during the announcement period and becomes negative at the moment of construction (Figure 8.4). Finally, the positive effect associated with the *De*

la Concorde Station may be due to its multimodal connection to the suburban train or the green landscaping developments.

Overall, even if the analysis is limited to single-family properties, the economic impacts of the transportation system remain subtle and doubts remain about the whole project's profitability at this time. The train enjoys a reputation making it the preferred choice when the time comes to develop a new public transport network. However, other alternatives, such as the bus, can be just as effective at lower costs if properly designed (Hensher, 2016).

8.2.8 Conclusion

The impact of a rail transit system development on property values has been well studied, generally leading to a positive effect on property transaction prices. However, the construction of such infrastructures is part of a long process and may fluctuate over space and time. The analysis aimed at decomposing the effect of proximity to a metro line extension per project implementation phase (*announcement, construction, operation*) and new stations. A temporal and spatial decomposition of the effect is compared to a more global approach often adopted by scholars (all stations and periods combined). Considering the extension of the Montréal subway to the city of Laval, a spatial difference-in-differences (SDID) approach based on a repeated-sales (RS) approach is applied using 5,422 pairs of transactions collected between 1995 and 2013 to decompose the spatial and temporal impact of the new subway line on single-family property values. The results highlighted the complexity of the assessment of a transportation service's impact on real estate values depending on temporal and spatial considerations.

From a global perspective (all stations combined), no significant effect is observed, neither after the announcement of the project nor during the construction period. Yet, considering the temporal aspect, the new transportation service positively influences property values after the first operation. The decomposition of the effect over the different stations shows non-homogenous effects over space. In fact, the estimated effect appears to be localised around a single station (*De la Concorde* station) during the construction and operation periods. Moreover, the measured effect is higher when decomposed per station. From a methodological perspective, the SDID estimator produces more precise estimations and allows considering spillover effect in the computation for a larger total effect.

Overall, the empirical application presented an innovative modelling approach to decompose the effect of a new metro infrastructure to answer some critics suggesting a lack of temporal (Yan et al., 2012) and spatial dimensions (Hess and Almeida, 2007) in such analyses. Using single-family properties, the results show temporally and spatially localised impacts that could relativize ambitious expectations associated with a mass transit system in an urban development perspective.

CONCLUSION GÉNÉRALE

RAPPEL DU CONTEXTE DE LA RECHERCHE

La Science régionale SR évolue au rythme de la notion d'espace (Claval, 2008) qui implique l'émergence de nouvelles problématiques d'analyse et de nouveaux défis auxquels les *regional scientists* doivent s'ajuster (Bolton and Jensen, 1995). L'une des critiques formulées à l'encontre de la SR porte sur ses méthodes d'analyse et de modélisation jugées trop rigides et inadéquates pour répondre aux problèmes du *vrai monde* (Bailly and Coffey, 1994, Coffey, 1992).

Dans les faits, si des réflexions sont nécessaires sur l'application des méthodes de modélisation, les méthodes économétriques ont su évoluer afin, notamment, d'intégrer la dimension spatiale dans les analyses. La démarche économétrique spatiale s'inspire de la première loi de la géographie de Tobler (1979) et postule que tous les éléments situés dans l'espace sont en relation. Elle suppose également que la force de cette relation fluctue en fonction de la distance. Les outils économétriques spatiaux sont développés afin de mesurer et maîtriser les effets de cette dépendance spatiale sur les estimations et leur application dépasse largement le cadre de l'analyse économique (Dubé and Devaux, 2014).

La dimension spatiale constitue donc un élément central dans l'analyse des phénomènes régionaux et territoriaux. Néanmoins, ces phénomènes sont également inscrits dans une logique dynamique. Pour cette raison, l'intégration croissante de la dimension temporelle à l'analyse économétrique spatiale (Arbia, 2011) représente une avancée majeure et permet de considérer les relations de causalité dans les analyses (Anselin, 2002).

Pourtant, lors de l'analyse de microdonnées non répétées et empilées dans le temps, la formalisation des processus de dépendance spatiotemporelle dans la démarche de modélisation fait souvent défaut. Même si des approches spatiotemporelles sont proposées (expansion de Casetti, modèle STAR, matrice spatiotemporelle générale), il est commun de ne pas considérer la dimension temporelle dans les analyses empiriques. En conséquence, les chercheurs postulent l'observation simultanée des données et supposent des relations spatiales incohérentes avec la logique temporelle (*The arrow of time*) (Thanos et al., 2016), surestimant ainsi les relations de dépendance.

L'enjeu est important dans la mesure où les estimations économétriques risquent d'être invalidées, tout comme les conclusions tirées de ces analyses. Si la présence d'autocorrélation spatiale viole l'hypothèse d'indépendance de la régression linéaire estimée par MCO et affecte les tests statistiques standards (Anselin and Griffith, 1988), elle doit nécessairement être prise en compte, mais de manière adéquate. La dimension spatiale n'est parfois pas la seule épaisseur que peut contenir une base de données; elle peut également intégrer la dimension temporelle. L'omission de cette dimension aurait ainsi pour conséquence de négliger les effets dynamiques, estimés par le biais des coefficients autorégressifs, et ainsi nuire à l'interprétation des effets marginaux totaux. En conséquence, les estimations et les conclusions de recherche peuvent être faussées, et ce même si la dimension spatiale est intégrée (Elhorst, 2010).

Sur la base de cas d'application propres aux ventes immobilières, un exemple flagrant de données spatiales non répétées et empilées dans le temps, la thèse comporte quatre articles abordant les questions de modélisation dans un contexte spatial et temporel particulier. La thèse est décomposée en deux sections. Par l'intermédiaire des deux premiers articles, elle vise d'abord à évaluer l'impact de l'omission de la dimension temporelle sur les statistiques spatiales lorsque des données en coupes transversales empilées sont utilisées en étudiant l'impact du processus générateur de données (PGD) sur

la détection et la correction de la dépendance spatiale dans certaines procédures économétriques. Les deux articles suivants proposent des applications empiriques misant sur une approche par ventes répétées, permettant de contrôler en large partie l'effet de variation spatiale pour se concentrer sur la dimension temporelle, et l'estimation d'un modèle DID (Bailey et al., 1963, Case and Shiller, 1987, Dubé et al., 2011b, Dubé et al., 2013b) afin d'étudier l'effet d'un changement de milieu par le biais des externalités liées au marché immobilier.

LA MÉTA-ANALYSE COMME STRATÉGIE DE RECHERCHE

Le premier article s'inscrit de manière complémentaire aux études précédemment entreprises par Smith et Wu (2009) ou encore Dubé et Legros (2013b, 2014c). L'article a été publié dans la revue *Journal of Real Estate Literature* et a notamment fait l'objet d'une présentation au colloque annuel de l'*American Real Estate Society* (San Diego, 2014).

Contrairement aux travaux précurseurs, l'approche préconisée valorise les outils d'analyse proposés par la méta-analyse (MA). Cette démarche permet d'intégrer et comparer un nombre important d'études de cas ayant recours à la modélisation hédonique en immobilier. Plusieurs variables de contrôle sont utilisées pour expliquer la variation dans l'indice I de Moran (détection de l'autocorrélation spatiale) et les coefficients autorégressifs ρ (modèle avec spécification autorégressive sur la variable dépendante- SAR) et λ (modèle avec spécification autorégressive sur les termes d'erreurs – SEM). De manière générale, les résultats suggèrent que les statistiques spatiales sont effectivement sensibles à la dimension temporelle, mais que les effets du PGD demeurent complexes. Il appert notamment que la spécification d'une matrice de pondérations spatiotemporelles n'affecte pas la mesure de l'autocorrélation spatiale ni le coefficient ρ associé à un processus autorégressif sur la variable dépendante. En revanche, le coefficient λ associé à processus autorégressif sur le

terme d'erreur s'en trouve diminué. De manière intéressante, les analyses suggèrent également que le problème de surconnectivité (Smith, 2009) pourrait ne pas se limiter à la proximité spatiale, mais s'étendre également à la proximité temporelle.

Si l'utilisation de la MA s'est révélée être une approche pertinente, l'analyse des données a mis en évidence certaines lacunes associées à la méthodologie employée. D'abord, alors que plus de 300 articles scientifiques ont été recensés lors de la revue de la littérature initiale, le nombre d'études retenues pour l'analyse s'est vu grandement diminué en raison d'informations manquantes dans la description méthodologique ou de l'absence des statistiques d'intérêt et de leur variance. L'approche se limite à l'étude des stratégies de modélisation les plus employées et ne permet pas de comparaison avec des approches alternatives, ces dernières étant sous-représentées dans l'échantillon.

En conséquence, le second article propose une approche différente permettant de contrôler l'effet des choix méthodologiques dans les études de cas. Une approche méthodologique originale est proposée avec pour objectifs : i) d'obtenir toutes les statistiques souhaitées estimées à partir d'échantillons empiriques et ii) de permettre un meilleur contrôle de la méthodologie employée pour estimer ces statistiques. La méthode utilisée repose sur une démarche d'échantillonnage inspirée du *bootstrap* et a fait l'objet d'une présentation lors d'un séjour à l'université de Dijon. Une série de sous-échantillons aléatoires a été créée à partir d'une large base de données (transactions immobilières de Lucas County, OH (É.U.) collectées entre 1993 et 1998). Les caractéristiques de chaque échantillon ont été conservées pour être comprises comme variables explicatives lors de la méta-régression. Les statistiques spatiales souhaitées ont été estimées et calculées à partir de plusieurs PGD dont les caractéristiques ont également été conservées. Les résultats suggèrent que l'indice I de Moran n'est pas significativement influencé par l'inclusion ou l'omission de la dimension temporelle contrairement au coefficient autorégressif ρ qui y est plus sensible. Celui-ci est davantage influencé par les caractéristiques de l'échantillon ainsi

que les choix exogènes de modélisation lorsqu'une approche strictement spatiale (omission du temps) est employée. Dans le prolongement de la première MA, les résultats renvoient à nouveau au problème de surconnectivité mentionné par Smith (2009). Une approche évacuant la dimension temporelle génère des relations abusives entre les observations, nuisant ainsi à la mesure des processus dynamiques et de causalité. Il existe donc un risque de biais dans les estimations et, par conséquent, dans les résultats d'études. Les résultats de recherche ont par la suite donné lieu à des présentations à l'*American Association of Geographer* (AAG - Chicago, 2015) ainsi qu'à la *Western Regional Science Association* (WRSA - Tuscon, 2015). L'article a été soumis à la revue *Letters in Spatial Resource and Sciences*.

Les deux premiers articles de la thèse ont souligné la nécessité de contrôler les effets de dépendances spatiotemporelles dans les procédures économétriques appliquées aux microdonnées spatiales non répétées et empilées dans le temps. Les méta-analyses ont permis de mettre en évidence qu'une approche de modélisation strictement spatiale pouvait nuire aux estimations. L'une des raisons se trouve reliée aux difficultés posées par la modélisation explicite de l'espace. En effet, les éléments environnementaux sont susceptibles de fluctuer à travers le temps et l'espace. Leur intégration aux procédures économétriques suppose l'utilisation d'effets de localisation parfois imprécis et renvoyant aux problèmes d'agrégation (Anselin, 2002, Haining et al., 2009) et d'aires modifiables (Dubé and Brunelle, 2014, Openshaw, 1983). À cela s'ajoute la non-linéarité des relations spatiales (McMillen, 2010) ainsi que les problèmes associés à l'omission de variables (Le Gallo, 2002, Osland, 2013). Les méta-analyses ont également souligné l'influence de la forme fonctionnelle du modèle sur les statistiques spatiales.

DES APPLICATIONS EMPIRIQUES

Les deux articles suivants constituent des applications empiriques préconisant une approche par ventes répétées destinée à maîtriser une partie des processus de dépendances spatiale et temporelle.

La théorie hédonique introduite par Rosen (1974) permet de décomposer le prix de vente d'un bien complexe et d'estimer un prix implicite pour les caractéristiques qui le composent. L'une des attentes envers l'analyse régionale est l'évaluation de politiques publiques. Dans un contexte urbain, l'une des stratégies d'évaluation de ces impacts économiques passe par l'analyse des valeurs foncières : le marché immobilier étant reconnu pour être étroitement dépendant des facteurs de localisation. Les effets d'une politique peuvent donc être évalués selon les fluctuations des valeurs foncières. Un prix implicite associé à la proximité de certaines aménités ou services peut ainsi être estimé.

Dans une logique de développement urbain, il est généralement considéré qu'une augmentation des valeurs immobilières se traduit par une augmentation des revenus de taxe pour les municipalités. Ces dernières y voient un moyen d'assurer, au moins en partie, un retour sur un investissement dans l'aménagement ou l'implantation d'un service. L'application d'une méthodologie adéquate est primordiale pour une juste évaluation de ces impacts. Les deux derniers articles de la thèse sont inspirés de cette approche permettant au développement (urbain) de s'autofinancer en partie. La question demeure de voir si les projets choisis par les décideurs publics sont aptes à générer des retombées économiques, et dans quelle mesure ces impacts peuvent permettre un retour sur les investissements publics.

L'approche par ventes répétées (Case and Shiller, 1987) est utilisée afin de répondre à certaines des limites identifiées dans les approches plus classiques (deux premières MA). Elle vise à expliquer la variation du prix d'un bien donné entre sa vente et sa revente. Cette approche est bien connue pour le développement d'indices de prix retraçant l'évolution des

valeurs immobilières (Bailey et al., 1963). Elle a également l'avantage de contrôler efficacement les effets de la structure spatiale postulée constante à travers le temps (Dubé et al., 2011b, Dubé et al., 2013b). De plus, l'utilisation de ventes répétées permet de libérer le chercheur des problèmes associés à l'omission de variables (Dubé et al., 2014) ainsi qu'à la forme fonctionnelle (McMillen, 2010). Afin de considérer une partie de la dimension temporelle, un estimateur de type différence-en-différences (DID) est utilisé. Celui-ci permet d'isoler l'effet d'un changement dans l'aménagement sur les valeurs foncières.

Le troisième article vise à évaluer l'impact d'une politique d'aménagement dans le contexte d'un quartier historique. La restauration de la rue d'Auteuil, dans le Vieux-Québec, a été effectuée durant l'été 2005 et se distingue d'une politique d'aménagement urbain traditionnelle par sa dimension patrimoniale dont la recherche tente d'estimer l'effet sur les valeurs immobilières. L'effet de la politique de patrimonialisation est étudié à partir de deux définitions de la proximité basées sur la distance réseau et la distance euclidienne. Les résultats illustrent la complexité de l'impact d'une telle politique. En effet, l'effet mesuré constitue un équilibre entre des externalités positives (effets symbolique et idéologique, verdure, espace piéton maximisé) et des externalités négatives (revêtement de sol bruyant, places de stationnement modifiées, axe routier rétréci) causées par l'implémentation de la politique patrimoniale. De manière générale, les conclusions montrent que ces deux types d'effets s'annulent à proximité immédiate de la rue restaurée, alors que les externalités nettes proposent un impact négatif pour les condominiums situés sur les rues voisines. Ainsi, cet article présente une manière efficace de modéliser les effets spatiaux et de considérer une partie de la dimension temporelle. Pourtant, les effets dynamiques (*effets de débordements*) demeurent ignorés, ce qui constitue une lacune potentielle de l'approche. L'article a fait l'objet d'une présentation au colloque annuel de l'*American Real Estate Society* (ARES - Fort Myers, 2015) et a été soumis à la revue *Urban Geography*.

Le quatrième article propose de remédier à une des lacunes identifiées dans l'article précédent, soit l'omission des effets de débordements spatiaux dans un contexte temporel, tout en permettant aux effets de varier dans l'espace et dans le temps. La recherche vise à évaluer l'impact économique, par le biais de valeurs foncières, de l'extension du métro de Montréal vers la ville de Laval. Le développement ou l'extension d'un réseau de transport en commun est reconnu pour affecter la structure urbaine (Debrezion et al., 2007) et le prix du sol. Les valeurs foncières constituent un bon indice de la mobilité (Burgess, 2008) et permettent donc d'en évaluer les impacts. L'article possède deux objectifs principaux : i) proposer une décomposition de l'effet mesuré selon les dimensions temporelle et spatiale plus détaillées et ii) intégrer au modèle DID les effets spatiaux de débordements influençant les effets marginaux mesurés dans le but d'accroître la validité des estimations. Puisqu'un tel projet de développement est généralement long et coûteux, il convient de décomposer l'impact en fonction des phases d'implantation du service (Agostini and Palmucci, 2008, Atkinson-Palombo, 2010, McDonald and Osuji, 1995, McMillen and McDonald, 2004) en plus de tenir compte des spécificités propres aux différentes stations (Hess and Almeida, 2007). L'étude empirique repose sur l'intégration de la dimension spatiale dans un estimateur de type DID (SDID) afin de décomposer l'effet total en effets direct et indirect. Les résultats montrent qu'une mesure de l'impact de la proximité au métro par période d'implantation du service (annonce, construction, mise en service) ne permet pas l'identification d'effets de spéculation après l'annonce publique du projet ou d'effets d'anticipation durant la période de construction. En revanche, l'ouverture au public du service présente un impact positif. La décomposition de l'effet par station montre finalement que seule la station *De la Concorde* présente un impact positif. Au-delà de l'effet du service, un effet d'anticipation est également observé. Ainsi, la mesure de l'effet global du nouveau service de transport est en réalité associée à une seule station. L'article a été amorcé lors d'un stage de recherche sous la supervision du professeur Philippe Apparicio à l'INRS Urbanisation (Montréal). Il a fait l'objet d'une présentation au colloque

annuel de la *Mid-continental Regional Science Association* (Saint-Louis, 2015) et a été soumis à la revue *Transportation Research Part A*.

EN RÉSUMÉ

Au final, la thèse était destinée à vérifier l'effet de la dimension temporelle lorsque des microdonnées spatiales non répétées et empilées dans le temps sont impliquées dans l'analyse des phénomènes régionaux. Bien au-delà de l'analyse immobilière et de la perspective économique, la disponibilité croissante de telles données géoréférencées ouvre la porte à de nouvelles problématiques de recherche associées à la localisation des firmes ou à leur survie, à l'analyse de politiques publiques ou au développement urbain, etc. Certaines réflexions méthodologiques s'imposent face l'application mécanique et peu appropriée des méthodes développées dans le cadre d'observations répétées.

La notion de dépendance spatiale est un concept devenu incontournable dans l'analyse régionale : « No other concept in empirical spatial research is as central to model building as is spatial autocorrelation » (Getis, 2008: 299). La recherche a cependant contribué à mettre en évidence que les observations sont certes situées dans l'espace, mais le sont également dans le temps. Les phénomènes régionaux et territoriaux sont inscrits dans des logiques dynamiques nécessitant l'intégration simultanée de ces dimensions dans les analyses économétriques spatiales.

De manière générale, les conclusions invitent à considérer la dimension temporelle au risque d'invalider les résultats d'études. La portée de ces résultats s'étend à tous les champs de recherche régionale, urbaine, rurale ou territoriale impliquant l'analyse quantitative de microdonnées non répétées. Si les individus sont effectivement au cœur de la SR

(Hägerstraand, 1970, Isard, 1960), les efforts pour le développement de la modélisation spatiotemporelle devraient être un chantier de recherche de première importance.

RÉFÉRENCES BIBLIOGRAPHIQUES

- AGOSTINI, C. A. & PALMUCCI, G. A. 2008. The anticipated capitalisation effect of a new metro line on housing prices. *Fiscal Studies*, 29, 233-256.
- AHLFELDT, G. M. & MAENNIG, W. 2010. Substitutability and Complementarity of Urban Amenities: External Effects of Built Heritage in Berlin. *Real Estate Economics*, 38, 285-323.
- AL-MOSAIND, M. A., DUEKER, K. J. & STRATHMAN, J. G. 1993. *Light-rail transit stations and property values: a hedonic price approach*.
- ALBERTO, I., BEAMONTE, A., GARGALLO, P., MATEO, P. M. & SALVADOR, M. 2010. Variable selection in STAR models with neighbourhood effects using genetic algorithms. *Journal of Forecasting*, 29, 728-750.
- ALONSO, W. 1964. *Location and land use. Toward a general theory of land rent*, Cambridge, MA, Harvard University Press.
- AMÉDÉE-MANESME, C.-O., BARONI, M., BARTHÉLÉMY, F. & DES ROSIERS, F. 2016. Market heterogeneity and the determinants of Paris apartment prices: A quantile regression approach. *Urban Studies*, 0042098016665955.
- ANDERSSON, H., JONSSON, L. & ÖGREN, M. 2010. Property Prices and Exposure to Multiple Noise Sources: Hedonic Regression with Road and Railway Noise. *Environmental and Resource Economics*, 45, 73-89.
- ANSELIN, L. 1980. Estimation methods for spatial autoregressive structures. *Regional Science Dissertation & Monograph Series, Cornell University*.
- ANSELIN, L. 1988. *Spatial econometrics: methods and models*, Springer.
- ANSELIN, L. 1990. Spatial dependence and spatial structure instability in applied regression analysis. *J. Reg. Sci.* 30, 185–207. *Journal of Regional Science*, 30, 185-207.
- ANSELIN, L. 1998. GIS research infrastructure for spatial analysis of real estate markets. *Journal of Housing Research*, 9, 113-133.
- ANSELIN, L. 2002. Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, 27, 247-267.

- ANSELIN, L. 2010. Thirty years of spatial econometrics. *Papers in Regional Science*, 89, 3-25.
- ANSELIN, L. & ARRIBAS-BEL, D. 2011. Spatial fixed effects and spatial dependence. GeoDa Center for Geospatial Analysis and Computation.
- ANSELIN, L. & BERA, A. K. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics Textbooks and Monographs*, 155, 237-290.
- ANSELIN, L., FLORAX, R. & REY, S. J. 2004. *Advances in spatial econometrics: methodology, tools and applications*, Springer.
- ANSELIN, L. & FLORAX, R. J. 1995. New directions in spatial econometrics.
- ANSELIN, L. & GRIFFITH, D. A. 1988. Do spatial effects really matter in regression analysis? *Papers in Regional Science*, 65, 11-34.
- ANSELIN, L. & LOZANO-GRACIA, N. 2008. Errors in variables and spatial effects in hedonic house price models of ambient air quality. *Empirical Economics*, 34, 5-34.
- ANSELIN, L. & REY, S. 1991. Properties of tests for spatial dependence in linear regression models. *Geographical analysis*, 23, 112-131.
- ARBIA, G. 2011. A Lustrum of SEA: Recent Research Trends Following the Creation of the Spatial Econometrics Association (2007-2011). *Spatial Economic Analysis*, 6, 377-395.
- ARMSTRONG JR, R. J. 1994. Impacts of commuter rail service as reflected in single-family residential property values. *Transportation Research Record*.
- ARMSTRONG, R. & RODRÍGUEZ, D. 2006. An Evaluation of the Accessibility Benefits of Commuter Rail in Eastern Massachusetts using Spatial Hedonic Price Functions. *Transportation*, 33, 21-43.
- ARNOTT, R. J. & MCMILLEN, D. P. 2008. *A companion to urban economics*, John Wiley & Sons.
- ASABERE, P. K., HUFFMAN, F. E. & MEHDIAN, S. 1994. The adverse impacts of local historic designation: the case of small apartment buildings in Philadelphia. *The Journal of Real Estate Finance and Economics*, 8, 225-234.
- ASABERE, P. K. H., FORREST E. 1997. Hierarchical zoning, incompatible uses and price discounts. *Real Estate Economics*, 25, 439.
- ATKINSON-PALOMBO, C. 2010. Comparing the capitalisation benefits of light-rail transit and overlay zoning for single-family houses and condos by neighbourhood type in metropolitan Phoenix, Arizona. *Urban studies*.

- AYDALOT, P. 1985. *Economie régionale et urbaine*, Editions Economica.
- BAE, C.-H. C., JUN, M.-J. & PARK, H. 2003. The impact of Seoul's subway Line 5 on residential property values. *Transport policy*, 10, 85-94.
- BAILEY, M. J., MUTH, R. F. & NOURSE, H. O. 1963. A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58, 933-942.
- BAILLY, A. S. 2009. La science régionale en perspective. *Géographie, économie, société*, 11, 1-7.
- BAILLY, A. S. & COFFEY, W. J. 1994. Regional science in crisis: A plea for more open and relevant approach. *Papers in Regional Science*, 73, 3-14.
- BAJIC, V. 1983. The effects of a new subway line on housing prices in metropolitan Toronto. *Urban Studies*, 20, 147-158.
- BAL, F. & NIJKAMP, P. 2001. In search of valid results in a complex economic environment: The potential of meta-analysis and value transfer. *European Journal of Operational Research*, 128, 364-384.
- BARANZINI, A. & RAMIREZ, J. V. 2005. Paying for quietness: the impact of noise on Geneva rents. *Urban studies*, 42, 633-646.
- BAUM-SNOW, N., KAHN, M. E. & VOITH, R. 2005. Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000 [with Comment]. *Brookings-Wharton Papers on Urban Affairs*, 147-206.
- BEAMONTE, A., GARGALLO, P. & SALVADOR, M. 2008. Bayesian inference in STAR models using neighbourhood effects. *Statistical Modelling*, 8, 285-311.
- BEAMONTE, A., GARGALLO, P. & SALVADOR, M. 2010a. Analysis of housing price by means of STAR models with neighbourhood effects: A Bayesian approach. *Journal of Geographical Systems*, 12, 227-240.
- BEAMONTE, A., GARGALLO, P. & SALVADOR, M. 2010b. Robust Bayesian inference in STAR models with neighbourhood effects. *Journal of Statistical Planning and Inference*, 140, 3047-3057.
- BEAMONTE, A., GARGALLO, P. & SALVADOR, M. J. 2013. Retrospective analysis of the dwelling price by means of STAR models with neighborhood effects. *Journal of Real Estate Research*, 35, 173-197.
- BECKMANN, M. J. 1976. A discourse on distance. *The Annals of Regional Science*, 10, 1-8.

- BELL, K. P. & BOCKSTAEL, N. E. 2000. Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data. *Review of Economics & Statistics*, 82, 72-82.
- BERTHOLD, E. 2012a. *Les quartiers historiques: pressions, enjeux, actions*, Presses de l'Université Laval.
- BERTHOLD, É. 2012b. *Patrimoine, Culture Et Mise en Récit: L'île D'Orléans Et la Place Royale de Québec*, Les Presses de l'Université Laval.
- BESAG, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 192-236.
- BESAG, J. 1975. Statistical analysis of non-lattice data. *The statistician*, 179-195.
- BHATTACHARJEE, A., CASTRO, E. & MARQUES, J. 2012. Spatial Interactions in Hedonic Pricing Models: The Urban Housing Market of Aveiro, Portugal. *Spatial Economic Analysis*, 7, 133-167.
- BIN, O., CRAWFORD, T. W., KRUSE, J. B. & LANDRY, C. E. 2008. Viewscapes and flood hazard: Coastal housing market response to amenities and risk. *Land Economics*, 84, 434-448.
- BIVAND, R. 1998. A review of spatial statistical techniques for location studies. Available: <http://www.nhh.no/geo/gib/gib1998/gib98-3/lund.html> (Accès le 20 octobre 2013).
- BLOMMESTEIN, H. J. 1983. Specification and estimation of spatial econometric models: A discussion of alternative strategies for spatial economic modelling. *Regional Science and Urban Economics*, 13, 251-270.
- BOLLINGER, C. R., IHLANFELDT, K. R. & BOWES, D. R. 1998. Spatial variation in office rents within the Atlanta region. *Urban Studies*, 35, 1097-1118.
- BOLTON, R. & JENSEN, R. C. 1995. Regional science and regional practice. *International Regional Science Review*, 18, 133-145.
- BORENSTEIN, M., HEDGES, L. V., HIGGINS, J. & ROTHSTEIN, H. R. 2009. *Introduction to Meta-analysis*, John Wiley and sons.
- BOUDEVILLE, J.-R., ANTOINE, S. & GUICHARD, O. 1968. *L'espace et les pôles de croissance: recherches et textes fondamentaux*, Presses universitaires de France Paris.
- BOURBONNAIS, R. 2003. *Econométrie, Ed, DUNOD*.
- BOWEN, W. M., MIKELBANK, B. A. & PRESTEGAARD, D. M. 2001. Theoretical and Empirical Considerations Regarding Space in Hedonic Housing Price Model Applications. *Growth & Change*, 32, 466.

- BOWES, D. R. & IHLANFELDT, K. R. 2001. Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*, 50, 1-25.
- BOXALL, P. C., CHAN, W. H. & MC MILLAN, M. L. 2005. The impact of oil and natural gas facilities on rural residential property values: a spatial hedonic analysis. *Resource and Energy Economics*, 27, 248-269.
- BOYCE, D. 2004. A short history of the field of regional science. *Papers in Regional Science*, 83, 31-57.
- BRAMLEY, G., LEISHMAN, C. & WATKINS, D. 2008. Understanding Neighbourhood Housing Markets: Regional Context, Disequilibrium, Submarkets and Supply. *Housing Studies*, 23, 179-212.
- BRANDT, S. & MAENNIG, W. 2011. Road noise exposure and residential property prices: Evidence from Hamburg. *Transportation Research Part D: Transport and Environment*, 16, 23-30.
- BRANDT, S. & MAENNIG, W. 2012a. The impact of rail access on condominium prices in Hamburg. *Transportation*, 39, 997-1017.
- BRANDT, S. & MAENNIG, W. 2012b. Perceived externalities of cell phone base stations: The case of property prices in Hamburg, Germany. *Environment and Planning A*, 44, 396-410.
- BRASINGTON, D. M. 2004. House Prices and the Structure of Local Government: An Application of Spatial Statistics. *Journal of Real Estate Finance & Economics*, 29, 211-231.
- BRASINGTON, D. M. & SARAMA JR, R. F. 2008. Deed types, house prices and mortgage interest rates. *Real Estate Economics*, 36, 587-610.
- BRUECKNER, J. K. 2001. Urban sprawl: Lessons from urban economics. *Brookings-Wharton papers on urban affairs*, 2001, 65-97.
- BURGESS, E. W. 2008. *The growth of the city: an introduction to a research project*, Springer.
- CALOZ, R. & COLLET, C. 2011. *Analyse spatiale de l'information géographique*, PPUR Presses polytechniques.
- CAN, A. 1990. The measurement of neighborhood dynamics in urban house prices. *Economic Geography*, 66, 254-272.
- CAN, A. 1992. Specification and estimation of hedonic housing price models. *Regional science and urban economics*, 22, 453-474.
- CAN, A. & MEGBOLUGBE, I. 1997. Spatial Dependence and House Price Index Construction. *Journal of Real Estate Finance and Economics*, 14, 203-222.

- CAPELLO, R. 2009a. Space, growth, and development. *Handbook of Regional Growth and Development Theories*. Cheltenham, UK: Edward Elgar.
- CAPELLO, R. 2009b. Spatial spillovers and regional growth: a cognitive approach. *European Planning Studies*, 17, 639-658.
- CAPOZZA, D. R., ISRAELSEN, R. D. & THOMSON, T. A. 2005. Appraisal, Agency and Atypicality: Evidence from Manufactured Homes. *Real Estate Economics*, 33, 509-537.
- CARRIÓN-FLORES, C. I., E. G. 2004. Determinants of residential land-use conversion and sprawl at the rural-urban fringe. *American Journal of Agricultural Economics*, 86, 889-904.
- CASE, B. P., H. O. WACHTER, S. M. 1997. Frequency of Transaction and House Price Modeling. *Journal of Real Estate Finance and Economics*, 14, 173-187.
- CASE, K. & SHILLER, R. 1989a. The Efficiency of the Market for Single-Family Homes. *American Economic Review*, 79, 125-37.
- CASE, K. E. & SHILLER, R. J. 1987. Prices of single family homes since 1970: New indexes for four cities. National Bureau of Economic Research Cambridge, Mass., USA.
- CASE, K. E. & SHILLER, R. J. 1989b. The Efficiency of the Market for Single-Family Homes. *The American Economic Review*, 125-137.
- CASETTI, E. 1972. Generating models by the expansion method: applications to geographical research. *Geographical analysis*, 4, 81-91.
- CASETTI, E. 1997. The expansion method, mathematical modeling, and spatial econometrics. *International regional science review*, 20, 9-33.
- CAVAILHÈS, J. 2005. Le prix des attributs du logement. *Economie et statistique*, 91-123.
- CAVAILHÈS, J., BROSSARD, T., FOLTÊTE, J.-C., HILAL, M., JOLY, D., TOURNEUX, F.-P., TRITZ, C. & WAVRESKY, P. 2009. GIS-Based Hedonic Pricing of Landscape. *Environmental and Resource Economics*, 44, 571-590.
- CERVERO, R. 2006. Effects of light and commuter rail transit on land prices: experiences in San Diego County. *University of California Transportation Center*.
- CERVERO, R. & DUNCAN, M. 2002. Benefits of proximity to rail on housing markets: Experiences in Santa Clara County. *Journal of Public Transportation*, 5.
- CERVERO, R. & DUNCAN, M. 2004. Neighbourhood composition and residential land prices: does exclusion raise or lower values? *Urban Studies*, 41, 299-315.

- CERVERO, R. & LANDIS, J. 1993. Assessing the impacts of urban rail transit on local real estate markets using quasi-experimental comparisons. *Transportation Research Part A: Policy and Practice*, 27, 13-22.
- CERVERO, R. & LANDIS, J. 1997. Twenty years of the Bay Area Rapid Transit system: Land use and development impacts. *Transportation Research Part A: Policy and Practice*, 31, 309-333.
- CHASCO, C. Y. & LÓPEZ, F. A. H. 2008. Is spatial dependence an instantaneous effect?: some evidence in economic series of Spanish provinces. *Estadística española*, 50, 101-118.
- CHEN, H., RUFOLO, A. & DUEKER, K. J. 1997. Measuring the impact of light rail systems on single family home values: a hedonic approach with GIS application. *Center for Urban Studies College of Urban and Public Affairs Portland State University: Available from: <http://www.upa.pdx.edu/CUS>*.
- CHEN, Y. 2012. On the four types of weight functions for spatial contiguity matrix. *Letters in Spatial and Resource Sciences*, 5, 65-72.
- CHEN, Y. 2013. New Approaches for Calculating Moran's Index of Spatial Autocorrelation. *PloS one*, 8, 1-14.
- CHRISTALLER, W. 1933. *Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*, University Microfilms.
- CLARK, D. E. & HERRIN, W. E. 1997. Historical preservation districts and home sale prices: evidence from the Sacramento housing market. *The Review of Regional Studies*, 27, 29-48.
- CLAVAL, P. 2008. Espace et territoire. *Géographie, économie, société*, 10, 157-184.
- CLIFF, A. D. & ORD, J. K. 1968. The problem of spatial autocorrelation. *Scott, A.J. (Ed.), London papers in regional science*. London: Pion.
- CLIFF, A. D. & ORD, J. K. 1973. *Spatial Autocorrelation*, London, Pion.
- COFFEY, W. 1992. Géographie, économie, science régionale. *Encyclopédie de géographie*, 159-76.
- COFFIN, D. A. 1989. The impact of historic districts on residential property values. *Eastern Economic Journal*, 15, 221-228.
- COHEN, J. P. & COUGHLIN, C. C. 2008. Spatial hedonic models of airport noise, proximity, and housing prices. *Journal of Regional Science*, 48, 859-878.

- CONWAY, D., LI, C. Q., WOLCH, J., KAHLE, C. & JERRETT, M. 2010. A Spatial Autocorrelation Approach for Examining the Effects of Urban Greenspace on Residential Property Values. *Journal of Real Estate Finance & Economics*, 41, 150-169.
- COULSON, N. E. & LEICHENKO, R. M. 2001. The internal and external impact of historical designation on property values. *The Journal of Real Estate Finance and Economics*, 23, 113-124.
- COURLET, C. 2001. *Territoires et régions, les grands oubliés du développement économique*, Editions L'Harmattan.
- CRESPO, R. & GRÊT-REGAMEY, A. 2013. Local hedonic house-price modelling for urban planners: advantages of using local regression techniques. *Environment and Planning B: Planning and Design*, 40, 664-682.
- CRESSIE, N. & WIKLE, C. K. 2011. *Statistics for spatio-temporal data*, John Wiley & Sons.
- DEBREU, G. 1959. *Theory of value: An axiomatic analysis of economic equilibrium*, Yale University Press.
- DEBREZION, G., PELS, E. & RIETVELD, P. 2007. The Impact of Railway Stations on Residential and Commercial Property Value: A Meta-analysis. *Journal of Real Estate Finance & Economics*, 35, 161-180.
- DEBREZION, G., PELS, E. & RIETVELD, P. 2011. The impact of rail transport on real estate prices: An empirical analysis of the Dutch housing market. *Urban Studies*, 48, 997-1015.
- DEKKERS, J. E. C. & VAN DER STRAATEN, J. W. 2009. Monetary valuation of aircraft noise: A hedonic analysis around Amsterdam airport. *Ecological Economics*, 68, 2850-2858.
- DERMISI, S. & MCDONALD, J. 2011. Effect of "green" (LEED and ENERGY STAR) designation on prices/sf and transaction frequency: The Chicago office market. *Journal of Real Estate Portfolio Management*, 17, 39-52.
- DERMISI, S. V. & MCDONALD, J. F. 2010. Selling Prices/Sq. Ft. of Office Buildings in Down town Chicago—How Much Is It Worth to Be an Old But Class A Building? *Journal of Real Estate Research*, 32, 1-21.
- DES ROSIERS, F. 1991. RESIVALU: an hedonic residential price model for the Quebec region, 1986-87. *Property Tax Journal*, 10, 227-55.
- DES ROSIERS, F. 2002. Power Lines, Visual Encumbrance and House Values: A Microspatial Approach to Impact Measurement. *Journal of Real Estate Research*, 23, 275.

- DES ROSIERS, F., THÉRIAULT, M., VOISIN, M. & DUBE, J. 2010. Does an Improved Urban Bus Service Affect House Values? *International Journal of Sustainable Transportation*, 4, 321-346.
- DEVAUX, N. & DUBE, J. 2016. About the influence of time on spatial dependence: A meta-analysis using real estate hedonic pricing models. *Journal of Real Estate Literature*, 24, 31.
- DEWEES, D. N. 1976. The effect of a subway on residential property values in Toronto. *Journal of Urban Economics*, 3, 357-369.
- DIAZ, R. B. & MCLEAN, V. Impacts of rail transit on property values. American Public Transit Association Rapid Transit Conference Proceedings, 1999.
- DONOVAN, G. H., CHAMP, P. A. & BUTRY, D. T. 2007. Wildfire risk and housing prices: A case study from Colorado Springs. *Land; Economics*, 83, 217-233.
- DORMONT, B. 2007. Introduction à l'économétrie. 518 pages.
- DUBÉ, J. 2010. *Vers une approche spatio-temporelle pour la modélisation des prix hédoniques: Une application à la région de Québec, 1986-2004*. Université Laval.
- DUBÉ, J., BAUMONT, C. & LEGROS, D. 2013a. Matrices de pondérations et contexte spatio-temporel en économétrie spatiale. *Canadian Journal of Regional Science/Revue Canadienne des Sciences Régionales*, 36, 55-71.
- DUBÉ, J. & BRUNELLE, C. 2014. Dots to dots: a general methodology to build local indicators using spatial micro-data. *The Annals of Regional Science*, 53, 245-272.
- DUBÉ, J., DES ROSIERS, F. & THÉRIAULT, M. 2011a. Segmentation spatiale et choix de la forme fonctionnelle en modélisation hédonique. *Revue d'Économie Régionale & Urbaine*, 9-37.
- DUBÉ, J., DES ROSIERS, F., THÉRIAULT, M. & DIB, P. 2011b. Economic impact of a supply change in mass transit in urban areas: A Canadian example. *Transportation Research Part A: Policy and Practice*, 45, 46-62.
- DUBÉ, J. & DEVAUX, N. 2014. L'économétrie spatiale au service de l'analyse territoriale: une approche mixte pour un système complexe. *Sciences du territoire, Tome 2: Défis méthodologiques*. Québec: Presses de l'Université du Québec.
- DUBÉ, J. & LEGROS, D. 2013a. Dealing with spatial data pooled over time in statistical models. *Letters in Spatial and Resource Sciences*, 6, 1-18.
- DUBÉ, J. & LEGROS, D. 2013b. A spatio-temporal measure of spatial dependence: An example using real estate data. *Papers in Regional Science*, 92, 19-30.

- DUBÉ, J. & LEGROS, D. 2014a. *Économétrie spatiale appliquée des microdonnées*, London, ISTE editions.
- DUBÉ, J. & LEGROS, D. 2014b. Spatial econometrics and spatial data pooled over time: towards an adapted modeling approach. *Journal of Real Estate Literature*, 22, 103-125.
- DUBÉ, J. & LEGROS, D. 2014c. Spatial econometrics and the hedonic pricing model: what about the temporal dimension? *Journal of Property Research*, 31, 333-359.
- DUBÉ, J. & LEGROS, D. 2015. Modeling Spatial Data Pooled over Time: Schematic Representation and Monte Carlo Evidences. *Theoretical Economics Letters*, 5, 132.
- DUBÉ, J., LEGROS, D., THÉRIAULT, M. & DES ROSIERS, F. 2014. A spatial Difference-in-Differences estimator to evaluate the effect of change in public mass transit systems on house prices. *Transportation Research Part B: Methodological*, 64, 24-40.
- DUBÉ, J., THÉRIAULT, M. & DES ROSIERS, F. 2012. Using a Fourier polynomial expansion to generate a spatial predictor. *International Journal of Housing Markets and Analysis*, 5, 177-195.
- DUBÉ, J., THÉRIAULT, M. & DES ROSIERS, F. 2013b. Commuter rail accessibility and house values: The case of the Montreal South Shore, Canada, 1992–2009. *Transportation Research Part A: Policy and Practice*, 54, 49-66.
- DUBIN, R. A. 1998. Predicting house prices using multiple listings data. *The Journal of Real Estate Finance and Economics*, 17, 35-59.
- DURANTON, G. 1997. La nouvelle économie géographique: agglomération et dispersion. *Économie & prévision*, 131, 1-24.
- DURANTON, G. 1999. Distance, land, and proximity: economic analysis and the evolution of cities. *Environment and Planning a*, 31, 2169-2188.
- DYETT, M. V., DOMBUSCH, D., FAJANS, M., FALCKE, C., GUSSMAN, V. & MERCHANT, J. 1979. *Land use and urban development impacts of BART: final report*, US Dept. of Transportation.
- ELHORST, J. P. 2001. Dynamic models in space and time. *Geographical Analysis*, 33, 119-140.
- ELHORST, J. P. 2010. Applied spatial econometrics: raising the bar. *Spatial Economic Analysis*, 5, 9-28.
- ELHORST, J. P. 2014a. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, Springer.

- ELHORST, J. P. 2014b. Spatial panel data models. *Spatial Econometrics*. Springer.
- FAWCETT, C. B. 1919. *Provinces of England: a study of some geographical aspects of devolution*, Williams and Norgate.
- FERNÁNDEZ-AVILÉS, G., MÍNG, R. & MONTERO, J. M. 2012. Geostatistical Air Pollution Indexes spatial hedonic models: The case of Madrid, Spain. *Journal of Real Estate Research*, 34, 243-274.
- FINGLETON, B. 2003. Externalities, economic geography, and spatial econometrics: conceptual and modeling developments. *International Regional Science Review*, 26, 197-207.
- FISHER, R. A. 1935. The design of experiments.
- FLORAX, R. J., NIJKAMP, P. & WILLIS, K. K. G. 2002. *Comparative Environmental Economic Assessment*, Edward Elgar Publishing.
- FUJITA, M., KRUGMAN, P. R. & VENABLES, A. 2001. *The spatial economy: Cities, regions, and international trade*, MIT press.
- GARDES, F., DUNCAN, G. J., GAUBERT, P., GURGAND, M. & STARZEC, C. 2005. Panel and Pseudo-Panel Estimation of Cross-Sectional and Time Series Elasticities of Food Consumption: The Case of U.S. and Polish Data. *Journal of Business & Economic Statistics*, 23, 242-253.
- GATZLAFF, D. H. & HAURIN, D. R. 1997. Sample selection bias and repeat-sales index estimates. *The Journal of Real Estate Finance and Economics*, 14, 33-50.
- GATZLAFF, D. H. & HAURIN, D. R. 1998. Sample Selection and Biases in Local House Value Indices. *Journal of Urban Economics*, 43, 199-222.
- GATZLAFF, D. H. & SMITH, M. T. 1993. The impact of the Miami Metrorail on the value of residences near station locations. *Land Economics*, 54-66.
- GELFAND, A. E., ECKER, M. D., KNIGHT, J. R. & SIRMANS, C. F. 2004. The Dynamics of Location in Home Price. *Journal of Real Estate Finance & Economics*, 29, 149-166.
- GELFAND, A. E., KIM, H.-J., SIRMANS, C. & BANERJEE, S. 2003. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98, 387-396.
- GEOGHEGAN, J., LYNCH, L. & BUCHOLTZ, S. 2003. Capitalization of open spaces into housing values and the residential property tax revenue impacts of agricultural easement programs. *Agricultural and Resource Economics Review*, 32, 33-45.

- GETIS, A. 2007. Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, 37, 491-496.
- GETIS, A. 2008. A history of the concept of spatial autocorrelation: a geographer's perspective. *Geographical Analysis*, 40, 297-309.
- GETIS, A. 2009. Spatial Weights Matrices. *Geographical Analysis*, 41, 404-410.
- GHINAMO, M. L. 2012. Explaining the variation in the empirical estimates of academic knowledge spillovers. *Journal of Regional Science*, 52, 606-634.
- GIBBONS, S. & MACHIN, S. 2008. Valuing school quality, better transport, and lower crime: evidence from house prices. *Oxford Review of Economic Policy*, 24, 99-119.
- GIBBONS, S. & OVERMAN, H. G. 2012. Mostly pointless spatial econometrics? *Journal of Regional Science*, 52, 172-191.
- GLAESER, E. L., KAHN, M. E. & RAPPAPORT, J. 2008. Why do the poor live in cities? The role of public transportation. *Journal of Urban Economics*, 63, 1-24.
- GLASS, G. V. 1976. Primary, secondary, and meta-analysis of research. *Educational researcher*, 5, 3-8.
- GOODCHILD, M. F., ANSELIN, L., APPELBAUM, R. P. & HARTHORN, B. H. 2000. Toward spatially integrated social science. *International Regional Science Review*, 23, 139-159.
- GRAHAM, B., ASHWORTH, G. J. & TUNBRIDGE, J. E. 2000. A Geography of heritage: Power, culture and economy. *London: Arnold*.
- GREENE, W. H. 1997. *Econometric Analysis*. Inglewood Cliffs.
- GREFFE, X. 2003. La valorisation économique du patrimoine.
- GRIFFITH, D. A. 2012. Spatial statistics: A quantitative geographer's perspective. *Spatial Statistics*, 1, 3-15.
- GUIGOU, J.-L., AYDALOT, P. & HURIOT, J.-M. 1974. *Théorie économique et utilisation de l'espace*, Centre d'étude des techniques économiques modernes.
- GUILLAUME, M. 1980. *La politique du patrimoine*, Éditions Galilée.
- HÄGERSTRAAND, T. 1970. What about people in regional science? *Papers in regional science*, 24, 7-24.
- HAINING, R., LAW, J. & GRIFFITH, D. 2009. Modelling small area counts in the presence of overdispersion and spatial autocorrelation. *Computational Statistics & Data Analysis*, 53, 2923-2937.

- HAINING, R. P. 2009. Spatial autocorrelation and the quantitative revolution. *Geographical Analysis*, 41, 364-374.
- HAM, C., CHAMP, P. A., LOOMIS, J. B. & REICH, R. M. 2012. Accounting for Heterogeneity of Public Lands in Hedonic Property Models. *Land Economics*, 88, 444-456.
- HAM, Y. J., MADDISON, D. J. & ELLIOTT, R. J. R. 2013. The valuation of landfill disamenities in Birmingham. *Ecological Economics*, 85, 116-129.
- HANDY, S. 2005. Smart growth and the transportation-land use connection: What does the research tell us? *International Regional Science Review*, 28, 146-167.
- HANSZ, A. J. & HAYUNGA, D. K. 2012. Club good influence on residential transaction prices. *Journal of Real Estate Research*, 34, 549-576.
- HARBORD, R. M. & HIGGINS, J. 2008a. Meta-regression in Stata. *Meta*, 8, 493-519.
- HARBORD, R. M. & HIGGINS, J. P. 2008b. Meta-regression in Stata. *Meta*, 8, 493-519.
- HARTOG, F. 2014. *Régimes d'historicité. Présentisme et expériences du temps*, Seuil.
- HEINTZELMAN, M. D. & ALTIERI, J. A. 2013. Historic preservation: Preserving value? *The Journal of Real Estate Finance and Economics*, 46, 543-563.
- HELBICH, M., BRUNAUER, W., VAZ, E. & NIJKAMP, P. 2014. Spatial Heterogeneity in Hedonic House Price Models: The Case of Austria. *Urban Studies*, 51, 390-411.
- HENSHER, D. A. 2016. Why is Light Rail Starting to Dominate Bus Rapid Transit Yet Again? *Transport Reviews*, 36, 289-292.
- HERBERTSON, A. J. 1905. The major natural regions: an essay in systematic geography. *The Geographical Journal*, 25, 300-310.
- HESS, D. B. & ALMEIDA, T. M. 2007. Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. *Urban Studies*, 44, 1041-1068.
- HIGGINS, C. D. & KANAROGLOU, P. S. 2016a. Forty years of modelling rapid transit's land value uplift in North America: moving beyond the tip of the iceberg. *Transport Reviews*, 1-25.
- HIGGINS, C. D. & KANAROGLOU, P. S. 2016b. A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region. *Journal of Transport Geography*, 52, 61-72.
- HIGGINS, J. & THOMPSON, S. G. 2002a. Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21, 1539-1558.

- HIGGINS, J. P. & THOMPSON, S. G. 2002b. Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21, 1539-1558.
- HINDSLEY, P., HAMILTON, S. E. & MORGAN, O. A. 2013. Gulf Views: Toward a Better Understanding of Viewshed Scope in Hedonic Property Models. *The Journal of Real Estate Finance and Economics*, 47, 489-505.
- HITE, D., CHERN, W., HITZHUSEN, F. & RANDALL, A. 2001. Property-Value Impacts of an Environmental Disamenity: The Case of Landfills. *Journal of Real Estate Finance and Economics*, 22, 185-202.
- HOLMES, T. P., MURPHY, E. A. & BELL, K. P. 2006. Exotic forest insects and residential property values. *Agricultural and Resource Economics Review*, 35, 155.
- HOTELLING, H. 1929. Stability in competition. *Economic Journal*, 39, 41-57.
- HUANG, B., WU, B. & BARRY, M. 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24, 383-401.
- HUANG, H., MILLER, G. Y., SHERRICK, B. J. & GÓMEZ, M. I. 2006. Factors influencing Illinois farmland values. *American Journal of Agricultural Economics*, 88, 458-470.
- HUANG, J. C. 2010. Deriving benefit measures with higher precision: A study of economic values of air quality. *Land Economics*, 86, 727-745.
- HUDSON, R. 2007. Regions and regional uneven development forever? Some reflective comments upon theory and practice. *Regional Studies*, 41, 1149-1160.
- HUI, E. C. M., CHAU, C. K., PUN, L. & LAW, M. Y. 2007. Measuring the neighboring and environmental effects on residential property value: Using spatial weighting matrix. *Building and Environment*, 42, 2333-2343.
- HUI, E. C. M., LIANG, C., WANG, Z. & WANG, Y. 2015. The roles of developer's status and competitive intensity in presale pricing in a residential market: A study of the spatio-temporal model in Hangzhou, China. *Urban Studies*, 0042098015572317.
- HUNTER, J. E. & SCHMIDT, F. L. 1990. Methods of meta-analysis: Correcting error and bias in research findings.
- HURIOT, J.-M. 1994. *Von Thünen économie et espace*, Economica.
- IHLANFELDT, K. R. 2003. Rail transit and neighborhood crime: the case of Atlanta, Georgia. *Southern Economic Journal*, 273-294.

- ISARD, W. 1949. The general theory of location and space-economy. *The Quarterly Journal of Economics*, 476-506.
- ISARD, W. 1960. Methods of regional analysis: an introduction to regional science.
- KENNEDY, P. 2003. *A guide to econometrics*, MIT press.
- KESTENS, Y., THÉRIAULT, M. & DES ROSIERS, F. 2004. The impact of surrounding land use and vegetation on single-family house prices. *Environment and Planning B: Planning and Design*, 31, 539-567.
- KESTENS, Y., THÉRIAULT, M. & DES ROSIERS, F. 2006. Heterogeneity in hedonic modelling of house prices: looking at buyers' household profiles. *Journal of Geographical Systems*, 8, 61-96.
- KIM, C. W., PHIPPS, T. T. & ANSELIN, L. 2003. Measuring the benefits of air quality improvement: A spatial hedonic approach. *Journal of Environmental Economics and Management*, 45, 24-39.
- KIM, J. & GOLDSMITH, P. 2009. A Spatial Hedonic Approach to Assess the Impact of Swine Production on Residential Property Values. *Environmental and Resource Economics*, 42, 509-534.
- KIM, K. & LAHR, M. L. 2014. The impact of Hudson-Bergen Light Rail on residential property appreciation. *Papers in Regional Science*, 93, S79-S97.
- KLOOSTERMAN, R. C. & MUSTERD, S. 2001. The polycentric urban region: towards a research agenda. *Urban Studies*, 38, 623-634.
- KNAAP, G. J., DING, C. & HOPKINS, L. D. 2001. Do plans matter? The effects of light rail plans on land values in station areas. *Journal of Planning Education and Research*, 21, 32-39.
- KOSTOV, P. 2009. A spatial quantile regression hedonic model of agricultural land prices. *Spatial Economic Analysis*, 4, 53-72.
- KRYVOBOKOV, M. 2010. Is it worth identifying service employment (sub) centres when modelling apartment prices? *Journal of Property Research*, 27, 337-356.
- KUETHE, T. H. 2012. Spatial fragmentation and the value of residential housing. *Land Economics*, 88, 16-27.
- KUMINOFF, N. V. & JARRAH, A. S. 2010. A new approach to computing hedonic equilibria and investigating the properties of locational sorting models. *Journal of Urban Economics*, 67, 322-335.
- LANCASTER, K. J. 1966. A new approach to consumer theory. *Journal of Political Economy*, 74, 132-157.

- LANDIS, J., GUHATHAKURTA, S. & ZHANG, M. 1994. Capitalization of transit investments into single-family home prices: A comparative analysis of five California rail transit systems. *University of California Transportation Center*.
- LANDIS, J. & LOUZENHEISER, D. 1995. BART Access and Office Building Performance. *University of California Transportation Center*.
- LANDRY, C. E. & HINDSLEY, P. 2011. Valuing beach quality with hedonic property models. *Land Economics*, 87, 92-108.
- LAZRAC, F., NIJKAMP, P., RIETVELD, P. & ROUWENDAL, J. 2014. The market value of cultural heritage in urban areas: An application of spatial hedonic pricing. *Journal of Geographical Systems*, 16, 89-114.
- LE GALLO, J. 2002. Économétrie spatiale : l'autocorrélation spatiale dans les modèles de régression linéaire. *Economie et Prevision*, 155, 139-157.
- LEGENDRE, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74, 1659-1673.
- LEGUIZAMON, S. J. & ROSS, J. M. 2012. Revealed preference for relative status: Evidence from the housing market. *Journal of Housing Economics*, 21, 55-65.
- LEICHENKO, R. M., COULSON, N. E. & LISTOKIN, D. 2001. Historic preservation and residential property values: an analysis of Texas cities. *Urban Studies*, 38, 1973-1987.
- LEONTIEF, W. W. 1986. *Input-output economics*, Oxford University Press on Demand.
- LESAGE, J. & PACE, K. 2009a. *Introduction to spatial econometrics*, Boca Raton London New-York, CRC press.
- LESAGE, J. & PACE, R. K. 2009b. *Introduction to spatial econometrics*.
- LESAGE, J. P. 2014. What regional scientists need to know about spatial econometrics. *Available at SSRN 2420725*.
- LESAGE, J. P. & PACE, R. K. 2014a. The biggest myth in spatial econometrics. *Econometrics*, 2, 217-249.
- LESAGE, J. P. & PACE, R. K. 2014b. Interpreting spatial econometric models. *Handbook of Regional Science*. Springer.
- LEWIS-WORKMAN, S. & BROD, D. 1997. Measuring the neighborhood benefits of rail transit accessibility. *Transportation Research Record: Journal of the Transportation Research Board*, 147-153.
- LI, W. & SAPHORES, J.-D. 2012. A spatial hedonic analysis of the value of urban land cover in the multifamily housing market in Los Angeles, CA. *Urban Studies*, 49, 2597-2615.

- LIAO, W. C. & WANG, X. 2012. Hedonic house prices and spatial quantile regression. *Journal of Housing Economics*, 21, 16-27.
- LIU, X. 2013. Spatial and Temporal Dependence in House Price Prediction. *Journal of Real Estate Finance and Economics*, 47, 341-369.
- LÖCHL, M. & AXHAUSEN, K. W. 2010. Modelling hedonic residential rents for land use and transport simulation while considering spatial effects. *Journal of Transport and Land Use*, 3.
- LÖSCH, A. 1944. *Die räumliche ordnung der wirtschafft*, Verlag von Gustav Fischer.
- LOWENTHAL, D. 2015. *The Past is a Foreign Country-Revisited*, Cambridge University Press.
- MAIER, G. & TRIPPL, M. 2009. Location/allocation of regional growth. *Handbook of Regional Growth and Development Theories*. heltenham, UK: Edward Elgar.
- MALEYRE, I. 1997. L'approche hédonique des marchés immobiliers. *Etudes foncières*, 22-29.
- MARSHALL, A. 1890. *Principles of Economics*, London, Macmillan and co.
- MASON, R. 2005. Economics and historic preservation. In: INSTITUTION, T. B. (ed.). Washington, DC: Metropolitan Policy POrogram.
- MASSICOTTE, G. 2008. *Sciences du territoire: perspectives québécoises*, PUQ.
- MCCANN, P. 2007. Observational equivalence? Regional studies and regional science. *Regional Studies*, 41, 1209-1222.
- MCDONALD, J. F. & OSUJI, C. I. 1995. The effect of anticipated transportation improvement on residential land values. *Regional science and urban economics*, 25, 261-278.
- MCMILLEN, D. P. 2003. Spatial autocorrelation or model misspecification? *International Regional Science Review*, 26, 208-217.
- MCMILLEN, D. P. 2010. Issues in spatial data analysis. *Journal of Regional Science*, 50, 119-141.
- MCMILLEN, D. P. 2012. Repeat Sales as a Matching Estimator. *Real Estate Economics*, 40, 743-771.
- MCMILLEN, D. P. & MCDONALD, J. 2004. Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983–1999. *Real Estate Economics*, 32, 463-486.
- MELO, P. C., GRAHAM, D. J. & NOLAND, R. B. 2009. A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39, 332-342.

- MIHAESCU, O. & VOM HOFE, R. 2012. The Impact of Brownfields on Residential Property Values in Cincinnati, Ohio: A Spatial Hedonic Approach. *Journal of Regional Analysis & Policy*, 42, 223-236.
- MILITINO, A. F., UGARTE, M. D. & GARCIA-REINALDOS, L. 2004. Alternative Models for Describing Spatial Dependence among Dwelling Selling Prices. *Journal of Real Estate Finance & Economics*, 29, 193-209.
- MILLER, H. J. 2004. Tobler's first law and spatial analysis. *Annals of the Association of American Geographers*, 94, 284-289.
- MILLS, E. S. 1969. the value of urban land. In: PERLOFF, H. (ed.) *The quality of urban environment*. Baltimore, MA: Resources for future, inc.
- MOHAMMAD, S. I., GRAHAM, D. J., MELO, P. C. & ANDERSON, R. J. 2013. A meta-analysis of the impact of rail projects on land and property values. *Transportation Research Part A: Policy and Practice*, 50, 158-170.
- MOINE, A. 2006. Le territoire comme un système complexe: un concept opératoire pour l'aménagement et la géographie. *L'Espace géographique*, 115-132.
- MORAN, P. 1963. *L'analyse spatiale en science économique*, Paris, Cujas.
- MORAN, P. A. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10, 243-251.
- MORAN, P. A. 1950. A test for the serial independence of residuals. *Biometrika*, 178-181.
- MORISSET, L. K. 2009. *Des régimes d'authenticité: essai sur la mémoire patrimoniale*, PUQ.
- MORO, M., MAYOR, K., LYONS, S. & TOL, R. S. 2013. Does the housing market reflect cultural heritage? A case study of Greater Dublin. *Environment and Planning A*, 45, 2884-2903.
- MULLEY, C. & TSAI, C.-H. P. 2016. When and how much does new transport infrastructure add to property values? Evidence from the bus rapid transit system in Sydney, Australia. *Transport Policy*.
- MULLIGAN, G. F., PARTRIDGE, M. D. & CARRUTHERS, J. I. 2012. Central place theory and its reemergence in regional science. *The Annals of Regional Science*, 48, 405-431.
- MUTH, R. F. 1969. *Cities and housing: the spatial pattern of urban residential land use*, Chicago, University of Chicago press.

- NAPPI-CHOULET, I. & MAURY, T. P. 2011. A spatial and temporal autoregressive local estimation for the paris housing market. *Journal of Regional Science*, 51, 732-750.
- NAPPI-CHOULET, I. P., MAURY, T. P. 2009. A spatiotemporal autoregressive price index for the paris office property market. *Real Estate Economics*, 37, 305-340.
- NARWOLD, A., SANDY, J. & TU, C. 2008. Historic designation and residential property values. *International real estate review*, 11, 83-95.
- NEELAWALA, P., WILSON, C. & ATHUKORALA, W. 2013. The impact of mining and smelting activities on property values: A study of Mount Isa city, Queensland, Australia. *Australian Journal of Agricultural and Resource Economics*, 57, 60-78.
- NELSON, A. C. & MCCLESKEY, S. 1989. Influence of Elevated Transit Stations on Neighborhood House Values. *Atlanta: Georgia Institute of Technology*.
- NELSON, J. P. 2004. Meta-Analysis of Airport Noise and Hedonic Property Values: Problems and Prospects. *Journal of Transport Economics and Policy*, 38, 1-27.
- NELSON, J. P. 2010. Valuing rural recreation amenities: Hedonic prices for vacation rental houses at deep Creek Lake, Maryland. *Agricultural and Resource Economics Review*, 39, 485-504.
- NEWBURN, D. A., BERCK, P. & MERENLENDER, A. M. 2006. Habitat and open space at risk of land-use conversion: Targeting strategies for land conservation. *American Journal of Agricultural Economics*, 88, 28-42.
- NIJKAMP, P. 2004. *Urban dynamics and growth: advances in urban economics*, Elsevier.
- NOONAN, D. S. 2007. Finding an impact of preservation policies: Price effects of historic landmarks on attached homes in Chicago, 1990-1999. *Economic Development Quarterly*, 21, 17-33.
- O'SULLIVAN, A. 2007. *Urban economics*, McGraw-Hill/Irwin.
- OPENSHAW, S. 1983. *The modifiable areal unit problem*, Geo books Norwich.
- ORD, K. 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70, 120-126.
- OSLAND, L. 2010. An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, 32, 289-320.
- OSLAND, L. 2013. The importance of unobserved attributes in hedonic house price models. *International Journal of Housing Markets and Analysis*, 6, 63-78.

- OSLAND, L. & PRYCE, G. 2012. Housing Prices and Multiple Employment Nodes: Is the Relationship Nonmonotonic? *Housing Studies*, 27, 1182-1208.
- OSLAND, L., THORSEN, I. & GITLESEN, J. P. 2007. Housing price gradients in a region with one dominating center. *Journal of Real Estate Research*, 29, 321-346.
- PACE, R. B., RONALD SIRMANS, C. F. 1998. Spatial Statistics and Real Estate. *Journal of Real Estate Finance & Economics*, 17, 5-13.
- PACE, R. K., BARRY, R., CLAPP, J. M. & RODRIQUEZ, M. 1998. Spatiotemporal Autoregressive Models of Neighborhood Effects. *Journal of Real Estate Finance and Economics*, 17, 15-33.
- PACE, R. K., BARRY, R., GILLEY, O. W. & SIRMANS, C. F. 2000. A method for spatial-temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, 16, 229-246.
- PACE, R. K. & LESAGE, J. P. 2004. Spatial autoregressive local estimation. *Spatial econometrics and spatial statistics*, 31-51.
- PAELINCK, J. H., KLAASSEN, L. H., ANCOT, J., VERSTER, A. & WAGENAAR, S. 1979. *Spatial econometrics*, Saxon House Farnborough.
- PARENT, O. & VOM HOFE, R. 2013. Understanding the impact of trails on residential property values in the presence of spatial dependence. *The Annals of Regional Science*, 1-21.
- PATERSON, R. W. & BOYLE, K. J. 2002. Out of sight, out of mind? Using GIS to incorporate visibility in hedonic property value models. *Land Economics*, 78, 417-425.
- PATTON, M. & MCERLEAN, S. 2003. Spatial effects within the agricultural land market in Northern Ireland. *Journal of Agricultural Economics*, 54, 35-54.
- PERDOMO, J. A. 2011. A methodological proposal to estimate changes of residential property value: Case study developed in Bogotá. *Applied Economics Letters*, 18, 1577-1581.
- PERROUX, F. 1950. Economic space: theory and applications. *The Quarterly Journal of Economics*, 89-104.
- PERROUX, F. 1964. *L'Économie du XXe siècle*, Presses Universitaires de France.
- PIGOTT, T. 2012. *Advances in meta-analysis*, Springer Science & Business Media.
- PIKE, A. 2007. Editorial: Whither regional studies? : Taylor & Francis.

- PIRAS, G. & PRUCHA, I. R. 2014. On the finite sample properties of pre-test estimators of spatial models. *Regional Science and Urban Economics*, 46, 103-115.
- POLÈSE, M. 1994. *Economie urbaine et régionale: logique spatiale des mutations économiques*, Economica.
- POLÈSE, M. & SHEARMUR, R. 2004. Is distance really dead? Comparing industrial location patterns over time in Canada. *International Regional Science Review*, 27, 431-457.
- POLÈSE, M., SHEARMUR, R. & TERRAL, L. 2015. *Économie urbaine et régionale. Géographie économique et dynamique des territoires.*, Paris, Economica.
- POLÈSE, M. & SHEARMUR, R. G. 2005. *Économie urbaine et régionale: introduction à la géographie économique*, Economica.
- PONSARD, C. 1956. Note sur la localisation de la firme. *Revue économique*, 101-116.
- POULIOT, D. 2006. Une histoire du patrimoine en Occident. *XVIIIe-XXIe siècle: du*.
- PREDÖHL, A. 1928. The theory of location in its relation to general economics. *The journal of political economy*, 371-390.
- PRICE, J. I., MCCOLLUM, D. W. & BERRENS, R. P. 2010. Insect infestation and residential property values: A hedonic analysis of the mountain pine beetle epidemic. *Forest Policy and Economics*, 12, 415-422.
- RANGER, T. O. & HOBSBAWM, E. J. 1983. *The invention of tradition*, Cambridge University Press Cambridge.
- REES, J. 1999. Regional science: From crisis to opportunity. *Papers in Regional Science*, 78, 101-110.
- REY, S. J. & JANIKAS, M. V. 2005. Regional convergence, inequality, and space. *Journal of Economic Geography*, 5, 155-176.
- RICARDO, D. 1824. *Plan for the Establishment of a National Bank*, J. Murray.
- RICKMAN, D. S. 2009. Neighborhood historic preservation status and housing values in Oklahoma County, Oklahoma. *Journal of Regional Analysis and Policy*, 39, 99-108.
- RIZZO, I. & MIGNOSA, A. 2013. *Handbook on the economics of cultural heritage*, Edward Elgar Publishing.
- ROBINSON, E. A. G. 1931. *Structure of competitive industry*, Chicago, University of Chicago press.
- ROSEN, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *The journal of political economy*, 82, 34-55.

- ROSENAU, P. M. 1991. *Post-modernism and the social sciences: Insights, inroads, and intrusions*, Princeton University Press.
- ROY, S., BYRNE, J. & PICKERING, C. 2012. A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones. *Urban Forestry & Urban Greening*, 11, 351-363.
- RYAN, S. 1999. Property values and transportation facilities: finding the transportation-land use connection. *Journal of planning literature*, 13, 412-427.
- RYAN, S. 2005. The value of access to highways and light rail transit: evidence for industrial and office firms. *Urban studies*, 42, 751-764.
- RYBERG-WEBSTER, S. & KINAHAN, K. L. 2014. Historic preservation and urban revitalization in the twenty-first century. *Journal of Planning Literature*, 29, 119-139.
- SAINT-PIERRE, D. 2008. Identité et patrimoine... deux notions au cœur des interventions patrimoniales publiques du Québec et du Canada». *Stratégies identitaires de conservation et de valorisation du patrimoine, Paris: L'Harmattan*, 115-124.
- SALVI, M. 2003. Spatial estimation of the impact of airport noise on residential housing prices. *Zurich Cantonal Bank, Real Estate Unit*.
- SAMARASINGHE, O. & SHARP, B. 2010. Flood prone risk and amenity values: a spatial hedonic analysis. *Australian Journal of Agricultural and Resource Economics*, 54, 457-475.
- SANDER, H., POLASKY, S. & HAIGHT, R. G. 2010. The value of urban tree cover: A hedonic property price model in Ramsey and Dakota Counties, Minnesota, USA. *Ecological Economics*, 69, 1646-1656.
- SAPHORES, J.-D. & LI, W. 2012. Estimating the value of urban green areas: A hedonic pricing analysis of the single family housing market in Los Angeles, CA. *Landscape and Urban Planning*, 104, 373-387.
- SCHIPPER, Y., NIJKAMP, P. & RIETVELD, P. 2002. Thirty years of aircraft noise value studies: a meta-analysis. *Comparative Environmental Economic Assessment*. Edward Elgar Publishing.
- SEDGLEY, N. H., WILLIAMS, N. A. & DERRICK, F. W. 2008. The effect of educational test scores on house prices in a model with spatial dependence. *Journal of Housing Economics*, 17, 191-200.
- SEYA, H., YAMAGATA, Y. & TSUTSUMI, M. 2013. Automatic selection of a spatial weight matrix in spatial econometrics: Application to a spatial hedonic approach. *Regional Science and Urban Economics*, 43, 429-444.

- SHEPPARD, S. 1999. Chapter 41 Hedonic analysis of housing markets. In: PAUL, C. E., S. MILLS (ed.) *Handbook of Regional and Urban Economics*. Elsevier.
- SIRMANS, G. S., MACDONALD, L. & MACPHERSON, D. A. 2010. A meta-analysis of selling price and time-on-the-market. *Journal of Housing Research*, 19, 139-152.
- SMALL, K. A. & STEIMETZ, S. S. 2012. Spatial hedonics and the willingness to pay for residential amenities. *Journal of Regional Science*, 52, 635-647.
- SMITH, T. & WU, P. 2009. A spatio-temporal model of housing prices based on individual sales transactions over time. *Journal of Geographical Systems*, 11, 333-355.
- SMITH, T. E. 2009. Estimation bias in spatial models with strongly connected weight matrices. *Geographical Analysis*, 41, 307-332.
- SO, H., TSE, R. Y. & GANESAN, S. 1997. Estimating the influence of transport on house prices: evidence from Hong Kong. *Journal of Property Valuation and Investment*, 15, 40-47.
- STEIMETZ, S. S. 2010. Spatial multipliers in hedonic analysis: a comment on "spatial hedonic models of airport noise, proximity, and housing prices". *Journal of Regional Science*, 50, 995-998.
- STEPHAN, F. F. 1934. Sampling errors and interpretations of social data ordered in time and space. *Journal of the American Statistical Association*, 29, 165-166.
- STUDENT 1914. The elimination of spurious correlation due to position in time or space. *Biometrika*, 179-180.
- SUN, H., TU, Y. & YU, S.-M. 2005. A Spatio-Temporal Autoregressive Model for Multi-Unit Residential Market Analysis*. *The Journal of Real Estate Finance and Economics*, 31, 155-187.
- SUN, H. T., Y. YU, S. M. 2005. A spatio-temporal autoregressive model for multi-unit residential market analysis. *Journal of Real Estate Finance and Economics*, 31, 155-187.
- TELLIER, L.-N. 1993. *Économie spatiale: rationalité économique de l'espace habité*, Boucherville, Québec: G. Morin.
- THANOS, S., BRISTOW, A. L. & WARDMAN, M. R. 2012. Theoretically consistent temporal ordering specification in spatial hedonic pricing models applied to the valuation of aircraft noise. *Journal of Environmental Economics and Policy*, 1, 103-126.

- THANOS, S., BRISTOW, A. L. & WARDMAN, M. R. 2015. Residential sorting and environmental externalities: the case of nonlinearities and stigma in aviation noise values. *Journal of Regional Science*, 55, 468-490.
- THANOS, S., DUBÉ, J. & LEGROS, D. 2016. Putting time into space: the temporal coherence of spatial applications in the housing market. *Regional Science and Urban Economics*, 58, 78-88.
- THISSE, J.-F. 1997. L'oubli de l'espace dans la pensée économique. *Région et Développement*, 6, 13-39.
- THROSBY, D. 2010. *The economics of cultural policy*, Cambridge University Press.
- THÜNEN, J. V. 1826. Der isolierte Staat. *Beziehung auf Landwirtschaft und Nationalökonomie*.
- TOBLER, W. 1979. Cellular geography. *Philosophy in geography*. Springer.
- TSUTSUMI, M. & SEYA, H. 2009. Hedonic approaches based on spatial econometrics and spatial statistics: application to evaluation of project benefits. *Journal of geographical systems*, 11, 357-380.
- TU, Y., YU, S. M. & SUN, H. 2004. Transaction-based office price indexes: A spatiotemporal modeling approach. *Real Estate Economics*, 32, 297-328.
- ULLMAN, E. L. 1958. Regional development and the geography of concentration. *Papers in Regional Science*, 4, 179-198.
- VAN DUIJN, M. & ROUWENDAL, J. 2013. Cultural heritage and the location choice of dutch households in a residential sorting model. *Journal of Economic Geography*, 13, 473-500.
- VIDAL DE LA BLACHE, P., MARTONNE, E. D. & BINGHAM, M. T. 1926. Principles of human geography.
- VOISIN, M., DUBÉ, J., THÉRIAULT, M. & DES ROSIERS, F. 2010. Does administrative segmentation have a place in property analysis? Quebec City under the microscope. *Les découpages administratifs sont-ils pertinents en analyse immobilière? Le cas de Québec*, 54, 249-274.
- VOITH, R. 1991. Transportation, sorting and house values. *Real Estate Economics*, 19, 117-137.
- VOITH, R. 1993. Changing capitalization of CBD-oriented transportation systems: Evidence from Philadelphia, 1970–1988. *Journal of Urban Economics*, 33, 361-376.
- WALKER, J. K. 2016. Silence is Golden: Railroad Noise Pollution and Property Values. *The Review of Regional Studies*, 46, 75-89.

- WALLNER, R. 2012. Gis measures of residential property views. *Journal of Real Estate Literature*, 20, 225-244.
- WALSH, P. J., MILON, J. W. & SCROGIN, D. O. 2011. The spatial extent of water quality benefits in urban housing markets. *Land Economics*, 87, 628-644.
- WEBER, A. 1909. *Über den Standort der Industrien*, JCB Mohr.
- WEINSTEIN, B. L., CLOWER, T. L., MEANS, F., GAGE, L., PHARR, M., PETTIBON, G. & GILLIS, S. 2002. An assessment of the DART LRT on taxable property valuations and transit oriented development.
- WILHELMSSON, M. 2002. Spatial Models in Real Estate Economics. *Housing, Theory & Society*, 19, 92-101.
- WINGO, L. J. 1961. *Transportation and Urban Land Use*. Baltimore, MD: The Johns Hopkins Press.
- WOOLDRIDGE, J. 2012. *Introductory econometrics: A modern approach*, Cengage Learning.
- WOOLDRIDGE, J. M. 2010. *Econometric analysis of cross section and panel data*, MIT press.
- YAN, S., DELMELLE, E. & DUNCAN, M. 2012. The impact of a new light rail system on single-family property values in Charlotte, North Carolina. *Journal of Transport and Land Use*, 5.
- YANKAYA, U. & CELIK, H. 2004. Modeling the Impacts of Rail Transit Investment On the Values of Residential Property: a Hedonic Price Approach in the Case of Izmir Subway, Turkey. *Report Izmir Institute of Technology, Department of Civil Engineering*.
- YE, X. & REY, S. 2013. A framework for exploratory space-time analysis of economic data. *The Annals of Regional Science*, 50, 315-339.
- YU, T. H., CHO, S. H. & KIM, S. G. 2012. Assessing the Residential Property Tax Revenue Impact of a Shopping Center. *Journal of Real Estate Finance and Economics*, 45, 604-621.
- YULE, U. 1926. Why do we sometimes get nonsense-correlations between Time-Series?--a study in sampling and the nature of time-series. *Journal of the royal statistical society*, 89, 1-63.
- ZAHIROVIC-HERBERT, V. & CHATTERJEE, S. 2012. Historic preservation and residential property values: evidence from quantile regression. *Urban Studies*, 49, 369-382.

- ZHU, B., FÜSS, R. & ROTTKE, N. 2011. The Predictive Power of Anisotropic Spatial Correlation Modeling in Housing Prices. *Journal of Real Estate Finance & Economics*, 42, 542-565.
- ZIMMERMANN, J.-B. 2008. Le territoire dans l'analyse économique : proximité géographique et proximité organisée. *REvue française de gestion*, 105-180.
- ZUKIN, S. 1996. *The cultures of cities*, Wiley-Blackwell.