

Documents de Travail

Laboratoire d'Économie de Dijon
Université de Bourgogne –
CNRS UMR 6307
Inserm U1200

Working Papers

La méta-analyse pour évaluer les stratégies de modélisation :
*Mesure de la dépendance spatiale pour des micro-données
collectées de manière non-répétée*

Nicolas DEVAUX – Diego LEGROS – Jean DUBE

[2015-10]

La méta-analyse pour évaluer les stratégies de modélisation: *Mesure de la dépendance spatiale pour des micro-données collectées de manière non-répétée*

Nicolas Devaux
Université du Québec à Rimouski, Québec (Canada)
nicolas.devaux@uqar.ca

Diego Legros
Université de Bourgogne, Dijon (France)

Jean Dubé
Université Laval, Québec (Canada)

Last update: October 7, 2015

Abstract

Le projet vise à recourir aux méthodes de la méta-analyse afin d'améliorer la compréhension des processus spatiotemporels qui structurent les données et d'évaluer des approches de modélisation alternatives. Plus précisément, la recherche tente de déterminer l'impact de l'omission de la dimension temporelle sur la détection de l'autocorrélation spatiale pour des micro-données collectées de manière non-répétées. Pour ce faire, une nouvelle stratégie d'échantillonnage est développée afin de dépasser les limites rencontrées dans une analyse basée sur une revue de la littérature. Le texte propose un bref retour sur l'évolution de l'économétrie spatiale jusqu'à l'inclusion croissante de la dimension temporelle et expose certaines approches de modélisation spatiotemporelles ainsi que leurs limites. L'analyse compare la spécification d'une matrice de pondérations strictement spatiales (S) à la spécification spatiotemporelle générale alternative (W) proposée par Smith et Wu (2009), Huang et al. (2010) et Dubé et Legros (2013a). L'impact des caractéristiques de l'échantillon sont aussi considérées. Les résultats suggèrent une sous-estimation de l'indice I de Moran et de sa variance ainsi qu'une influence significative des caractéristiques de l'échantillon lorsque la dimension temporelle est omise.

MOTS CLEFS: Méta-analyse; Modélisation spatiotemporelle; Autocorrélation spatiale.

1 Introduction

La production importante d'études de cas sur une thématique donnée génère une abondance de résultats parfois divergents. La méta-analyse (MA) est destinée à réaliser une synthèse exhaustive de ces résultats et d'expliquer les variations 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).

La MA peut également être considérée comme un temps de pause afin d'aider à clarifier et synthétiser les connaissances relatives à un phénomène donné (Pigott, 2012). À titre d'exemple, l'analyse des valeurs foncières présente un nombre important d'études de cas évaluant l'impact de la localisation sur les prix de vente. Quelques travaux recourent à la MA pour expliquer les variations de résultats entre les études. Debrezion et al. (2007) considèrent par exemple l'impact des stations ferroviaires sur les propriétés résidentielles et commerciales. Nelson (2004); Schipper et al. (2002) portent leur attention sur les nuisances sonores aériennes tandis que Sirmans et al. (2010, 2006) ciblent plutôt le temps passé sur le marché et les caractéristiques des bien vendus. À notre connaissance, aucune étude ne tente d'expliquer les variations de résultats en fonction des méthodes de modélisation.

La dépendance spatiale est considérée dans l'analyse foncière depuis les années 90 (Dubin, 1998; Can, 1992) et il y a désormais un consensus sur l'importance de la localisation dans la détermination des ces valeurs¹. Néanmoins, il est reconnu que le marché immobilier intègre une dynamique spatiotemporelle (Huang et al., 2010; Smith and Wu, 2009) pourtant l'omission de la dépendance temporelle n'est pas rare dans les études empiriques. L'un des défis réside dans la conjugaison des relations de dépendance spatiales et temporelles pour des données collectées de manière sporadique. Une mauvaise interprétation de la structure spatiotemporelle pourrait affecter la détection et la correction de l'autocorrélation spatiale et par le fait même nuire aux résultats des méthodes économétriques spatiales. Il y a un besoin d'améliorer notre compréhension des processus spatiaux et spatiotemporels: «There is a need to better understand the fundamental processes behind the spatial and space-time correlation that is incorporated into our models» (Anselin, 2010, p.18). La recherche s'inscrit dans cette perspective et vise à réaliser une MA afin d'évaluer l'impact d'une approche strictement spatiale dans un contexte spatiotemporel. Dans la continuité d'une première MA (Devaux et Dubé, *à venir*), une nouvelle méthode d'échantillonnage est proposée afin d'en mesurer les conséquences sur la détection de la dépendance spatiale.

L'article est divisé en 4 sections. La première section discute de l'intégration de la dimension temporelle dans les méthodes économétriques spatiales afin de décrire la problématique associée aux micro-données spatiales collectées de manière non-répétées. Les principales approches de modélisation de ces données dans un contexte spatiotemporel et leurs limites sont ensuite présentées.

¹L'impact de l'accessibilité sur les valeurs foncières s'explique pour la théorie «bid-rent» (Alonso, 1964; Mills, 1969; Muth, 1969) qui émerge de la théorie économique urbaine traditionnelle selon laquelle les villes gravitent autour d'un centre d'affaire généralement nommé *central business district - CBD* (Ahlfeldt et Wendland, 2013). Des travaux plus récents tentent d'appréhender une structure polycentrique plus complexe des villes à travers des forces d'agglomération et de dispersion (idem). Selon la théorie *bid-rent*, l'impact de l'accessibilité sur la variation des valeurs foncières s'explique par une fonction décroissante du prix que le consommateur est prêt à payer selon la distance qui le sépare de certains points d'intérêts (O'Sullivan, 2003).

La section suivante présente la méthodologie de la recherche. La mesure de l'autocorrélation spatiale (indice I de Moran) y est d'abord présentée, puis la stratégie d'échantillonnage et de collecte de données sont décrites ainsi que la démarche de modélisation propre à la méta-analyse. La troisième section propose une discussion des résultats obtenus. Finalement, une conclusion propose une synthèse de ces constats.

2 Modélisation spatiotemporelle

2.1 Méthodes économétriques spatiales et micro-données

L'économétrie spatiale trouve ses fondements dans la première loi de la géographie de Tobler selon laquelle les objets géographiquement proches sont liés mais que la force de cette relation varie est fonction de leur proximité: «everything is related to everything else, but closer things more so» (Tobler, 1979, p. 236). Le phénomène de la *dépendance spatiale* a été mis en évidence par certains pionniers² des mesures de cet effet ont été développées (nommée autocorrélation spatiale³) ainsi qu'à son contrôle par l'usage de modèles auto-régressifs. Les méthodes économétriques spatiales ont pour leur part été développées pour répondre au problème d'autocorrélation des résidus du modèle de régression linéaire estimé par moindres carrés ordinaires. L'omission de la dépendance spatiale affecte les coefficients estimés et leur variance (LeSage and Pace, 2009; Pace, 2004; Griffith, 2005) et donc les résultats d'études.

L'indice de Moran (I) et le coefficient de Geary (c) sont les deux principales statistiques élaborées pour détecter la présence d'autocorrélation spatiale. Cliff and Ord (1973) établissent les propriétés de la distribution de ces deux statistiques qui permettront de cristalliser l'ère des tests d'hypothèses pour l'autocorrélation spatiale (Griffith, 2012, p.4) mais la statistique de Moran est la plus utilisée et la plus puissante statistiquement (Griffith, 2012). L'indice I de Moran a l'avantage d'avoir une interprétation simple similaire à celle du coefficient de corrélation (0 marquant l'absence d'autocorrélation). De plus, il serait moins sensible à une mauvaise spécification de la matrice de poids que d'autres et se présente comme un test de spécification général valable (Le Gallo, 2002; Anselin and Florax, 1995). Depuis, l'économétrie spatiale n'a pas fini d'évoluer au rythme de l'amélioration des méthodes statistiques⁴ jusqu'à intégrer de manière croissante la dimension temporelle (Arbia, 2011). Arbia 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). Pour leur part, Anselin et Griffith (1988) notent que si l'attention de l'économétrie a longtemps été portée sur les séries chronologiques, les possibles relations de dépendances entre les coupes transversales ont longtemps été ignorées. En considérant qu'une cause précède généralement un effet et qu'il existe une effet de proximité spatiale ou temporelle, les données spatiotemporelles ne devraient pas être considérées comme statistiquement indépendantes mais affectées par leur position

²Student, en 1914, serait le premier à en faire mention dans ses travaux (Le Gallo, 2002; Griffith, 2012). 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.

³Tel que proposé par Anselin et Rey (1991) nous distinguons 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.

⁴Plusieurs ouvrages méthodologiques de référence sont publiés⁵ et les méthodes qui y sont décrites sont appliquées dans une grande variété de disciplines⁶.

dans le temps et l'espace (Cressie and Wikle, 2011)

Ainsi, l'ajout de la dimension temporelle aux analyses spatiales se présente comme une étape incontournable. Les données en coupe transversale contiennent trop peu d'informations pour permettre l'identification de la nature précise des mécanismes de causalité sous-jacents (Gibbons and Overman, 2012). Il est alors impossible d'identifier si, par exemple, un *cluster* origine d'un processus de diffusion ou d'une hétérogénéité spatiale ou structurelle (Anselin, 2002). Les méthodes d'analyse de panel et de panel non balancé ont été largement abordées par la littérature (Elhorst, 2014b,a; Anselin et al., 2006; Baltagi et al., 2003, 2005) mais supposent l'utilisation de données collectées de manière répétée dont l'acquisition peut être difficile, voir impossible dans certains cas. Les micro-données peuvent être confrontées à cette difficulté (individus, firmes, logements, etc.). À l'ère du *Big Data* et de la disponibilité croissante de données géolocalisées, une réflexion méthodologique semble s'imposer pour analyser des données inscrites dans un contexte spatiotemporel en dehors d'une approche de panel : *Quel est l'impact de l'omission de la dimension temporelle sur les statistiques spatiales utilisant des données collectées de manière non-répétée?*

Les micro-données se caractérisent par leur état désagrégé ainsi que par la quantité d'information qu'elles véhiculent, particulièrement lorsqu'elles s'insèrent dans des dynamiques spatiotemporelles complexes. Lorsque les données sont collectées de manière continue au fil du temps, mais non répétées (observations différentes à chaque période temporelle), il peut être difficile de formaliser les relations de dépendance qui les structurent. Les valeurs résidentielles en constituent un exemple probant puisque chaque bien immobilier n'est pas vendu à intervalle de temps régulier:

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 et Wu, 2009 : 334)

Il est commun de constater l'omission de la dimension temporelle du processus qui structure les observations dans les études empiriques⁷. Le modélisateur suppose alors une réalisation simultanée des observations et, par conséquent, élimine les possibles contraintes de temps liées aux possibles relations entre les observations (Dubé and Legros, 2013b,a, 2014c). À ce jour, les données structurées en coupes transversales empilées sont généralement traitées avec les méthodes de modélisation strictement spatiales, ce qui suppose un effet nul du temps sur les phénomènes spatialisés. Si ce postulat est supposé vrai a priori (Dubé et al., 2013), il n'a que rarement été testé en pratique. Les conclusions de Dubé et Legros (2013) suggèrent d'ailleurs qu'une mauvaise interprétation de la structure des données ne permet pas de capter adéquatement l'autocorrélation. Les enjeux sont importants puisque la présence d'autocorrélation des résidus du modèle de régression estimé par moindres carrés ordinaires (MCO) affecte les estimateurs et leur variance (risque d'erreur de type 1) ainsi que la statistique d'ajustement du modèle (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). Dans le contexte de la science régionale, des conclusions erronées pourraient mener à l'implémentation de politiques publiques ou de développement contre-productives.

⁷Le tableau 6 présenté en annexe recense des études immobilières réalisées en coupe transversale.

2.2 Modélisation de micro-données collectées dans le temps et l'espace

Le défi réside dans la conjugaison des relations de dépendances spatiales et temporelles qui sont de nature différentes. Tandis que les observations s'inter-influencent de manière multidirectionnelle dans l'espace, la relation de dépendance est unidirectionnelle dans le temps de sorte que seules les observations passées influencent les observations à venir. La formalisation de ces relations passe par la spécifications de matrices de pondérations spatiales, temporelles et/ou spatiotemporelles⁸. Pierre angulaire de l'économétrie spatiale, la *juste* spécification de la matrice de pondérations fait débat. Pour certains, son caractère exogène en fait la principale faiblesse de l'économétrie spatiale (Bhattacharjee et al., 2012; Chen, 2012) tandis que d'autres suggèrent que son importance pourrait être relative ce qui en ferait le plus grand mythe des méthodes économétriques spatiales (LeSage and Pace, 2010).

Plusieurs approches sont proposées pour intégrer simultanément les relations spatiales et temporelles. D'aucuns optent pour une analyse en coupe transversale en incluant des effets fixes spatiaux et temporels, d'autres s'inspirent de l'analyse en panel et misent sur le pseudo-panel tandis que d'autres encore proposent des approches plus originales adaptées à la réalité des données spatiales collectées au fil du temps sans que les observations individuelles soient répétées (modèle STAR). Néanmoins, ces propositions méthodologiques peuvent présenter des limites auxquelles certains tentent de remédier avec la spécification d'une matrice de pondérations générale, W . Dans l'analyse immobilière, ces effets spatiaux et temporels peuvent être intégrés aux modèles statistiques par l'utilisation d'effets fixes associés à la période de transaction ou à la localisation du bien transigé. La construction de tels indices permet de mesurer les tendances du marché dans l'espace et le temps (Huang et al., 2010). Néanmoins ils reposent souvent sur l'agrégation spatiale (par quartiers, régions, etc.) ou temporelle (par mois, années, etc.) qui constitue des choix arbitraires. Considérant que chaque bien immobilier n'est pas vendu tous les ans, une approche par *ventes répétées* peut être utilisée mais diminue de manière importante la taille des échantillons.

Il est parfois proposé d'analyser ces données sous la forme d'un pseudo-panel (Deaton, 1985; Heckman and Robb, 1985; Moffitt, 1993), c'est à dire en agrégeant les observations par région qui, elles, sont géographiquement stables dans le temps. En d'autres termes, 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. Cette démarche permet de postuler des relations de dépendances spatiotemporelles mais engendre des pertes d'informations causées par l'agrégation et confronte le chercheur au problème des aires modifiables (*Modifiable Areal Unit Problem, MAUP* - Openshaw, 1983).

Comme le mentionne Haining (2009), 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: «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.» (ibid., p.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,

⁸Le lecteur intéressé pourra se référer à Dubé and Legros (2014a) pour plus de détails sur la spécification de ces matrices.

heteroskedasticity should be expected and an i.i.d. assumption for the errors is incompatible with aggregation rule» (Anselin, 2002, p.261).

Pace et al. (1998, 2000) proposent une approche autorégressive spatio-temporelle (modèle *STAR*) misant sur l'inclusion de *filtres* pour modéliser les effets de voisinages en respectant la chronologie de l'échantillon (propriétés voisines récemment transigées). 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). En assumant le processus autorégressif suivant $(I - W)Y = (I - W)X\beta + \varepsilon$, les auteurs proposent de construire un filtre basé sur la construction de matrices de pondérations spatiales (*S*), temporelles (*T*) et spatiotemporelles (*TS* et *ST*)⁹ tel que présenté à l'équation 1.

$$(I - \phi_T T - \phi_S S - \phi_{ST} ST - \phi_{TS} TS)Y = (I - \phi_T T - \phi_S S - \phi_{ST} ST - \phi_{TS} TS)X\beta + \varepsilon \quad (1)$$

Cette approche génère un ensemble de variables supplémentaires par le biais de la multiplication des matrices mais améliore le pouvoir explicatif du modèle et réduit significativement les erreurs de prédiction avec un nombre de variables de contrôle moindre (Liu, 2013). Ce modèle est à l'origine de plusieurs applications et développement (Nappi-Choulet, 2009; Nappi-Choulet and Maury, 2011; Beamonte et al., 2008, 2010b,a; Sun et al., 2005; Tu et al., 2004). Néanmoins, malgré l'ajout de matrices spatiotemporelles *ST* et *TS* le modèle *STAR*¹⁰ ne capte pas adéquatement les effets spatiotemporels (Smith and Wu, 2009; Dubé and Legros, 2013a). Smith et Wu (2009) montrent que les produits matriciels *ST* et *TS* proposés dans le modèle *STAR* ne permettent pas de capter adéquatement la nature indirect des processus spatiotemporels. En effet, deux observations spatialement ou temporellement distantes pourraient, par le jeu de relations intermédiaires indirectes, entretenir des relations. Telles que présenté dans le modèle 1, une observation *j* pourra influencer une observation *i* spatialement et temporellement proche mais pas les observations subséquentes. De plus, le modèle *STAR* corrige la présence d'autocorrélation de manière globale et assume une stationnarité des coefficients (Nappi-Choulet, 2009; Nappi-Choulet and Maury, 2011). Si la localisation est fixe, il est en revanche peu probable que les effets de l'autocorrélation spatiale soit constant dans le temps (Alberto et al., 2010).

Une dernière avenue, proposée par Dubé et Legros (2013a; 2013b) ou encore Smith et Wu (2009) et Huang et al. (2010) avant eux, suggère de considérer les données de transactions immobilières non répétées en coupes transversales empilées dans le temps (figure 1)¹¹.

Ils proposent la construction d'une matrice agrégée combinant les deux types de matrices (effet du temps et de l'espace) destinée à produire une pondération unique tout en tenant compte des relations de proximité et contraintes spatiales et temporelles (Dubé and Legros, 2014d). La matrice spatiotemporelle (*W*) est construite par le produit de Hadamard¹² des matrices de pondérations spatiales (*S*) et temporelles (*T*):

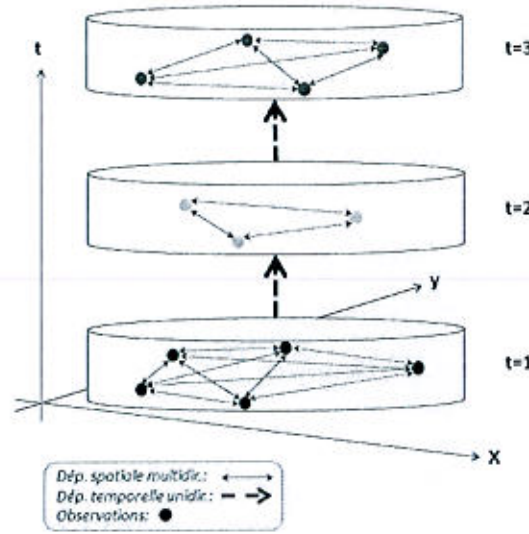
⁹Il est à noter que pour faciliter les calculs, les matrices *S* et *T* sont triangulaires inférieures. Les observations *y* sont organisées chronologiquement de sorte que l'observation la plus ancienne se trouve sur la première ligne. Ainsi, les éléments situés au-dessus de la diagonale de la matrice représentent le futur et sont donc nuls. Le produit des matrices *S* et *T* crée une matrice *ST* également triangulaire inférieure. Il en va de même pour la matrice *TS*.

¹⁰Les matrices *ST* et *TS* n'étant pas commutatives, les auteurs choisissent de les inclure toutes les deux: «it seems reasonable to allow the variables their own autoregressive parameters, ϕ_{ST}, ϕ_{TS} » (Pace et al., 1998, p.18).

¹¹Figure inspirée par Dubé et Legros, 2012

¹²Produit matriciel terme à terme

Figure 1: Observations en coupes transversales empilées - CTE



$$W_{(N_T \times N_T)} = S_{(N_T \times N_T)} \odot T_{(N_T \times N_T)} \tag{2}$$

$$W = \begin{pmatrix} 0 & w_{12} \times t_{12} & w_{13} \times t_{13} & \dots & w_{1N_T} \times t_{1N_T} \\ w_{21} \times t_{21} & 0 & w_{23} \times t_{23} & \dots & w_{2N_T} \times t_{2N_T} \\ w_{31} \times t_{31} & w_{32} \times t_{32} & 0 & \dots & w_{3N_T} \times t_{3N_T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{N_T1} \times t_{N_T1} & w_{N_T2} \times t_{N_T2} & w_{N_T3} \times t_{N_T3} & \dots & 0 \end{pmatrix} \tag{3}$$

Cette dernière proposition libère le chercheur des problèmes d'agrégation et des aires modifiables tout en respectant les logiques unidirectionnelles et multidirectionnelles de relations de dépendances temporelles et spatiales.

Huang et al. (2009) appliquent cette approche pour développer un modèle de régression géographiquement pondérée spatiotemporelle (GTWR) dont le pouvoir explicatif surpasse celui des approches stationnaire OLS. Smith et Wu (2010) ajoutent que contrairement aux produits matriciels de S et T proposé par le modèle STAR, le produit de Hadamard autorise les relations indirectes en préservant la triangularité inférieure de la matrice. Suite à des simulations et une application empirique, les auteurs concluent qu'une matrice basée sur le produit de Hadamard est adéquate à la modélisation de phénomènes spatiotemporels micro qui se réalisent de manière sporadique. Ils reconnaissent cependant que l'application sur des échantillon de taille restreinte pourraient bénéficier de procédures d'estimations plus robustes.

Dubé et Legros (2013a et 2013b) soulignent le rôle de cette spécification pour calculer des statistiques d'autocorrélation spatiales et les tests d'hypothèses. Les auteurs 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 et Legros, 2013a : 11). En comparant l'impact d'une première matrice exclusivement spatiale (S) et d'une seconde spatio-temporelle (W) sur la statistique de Moran, Dubé et Legros constatent une variation

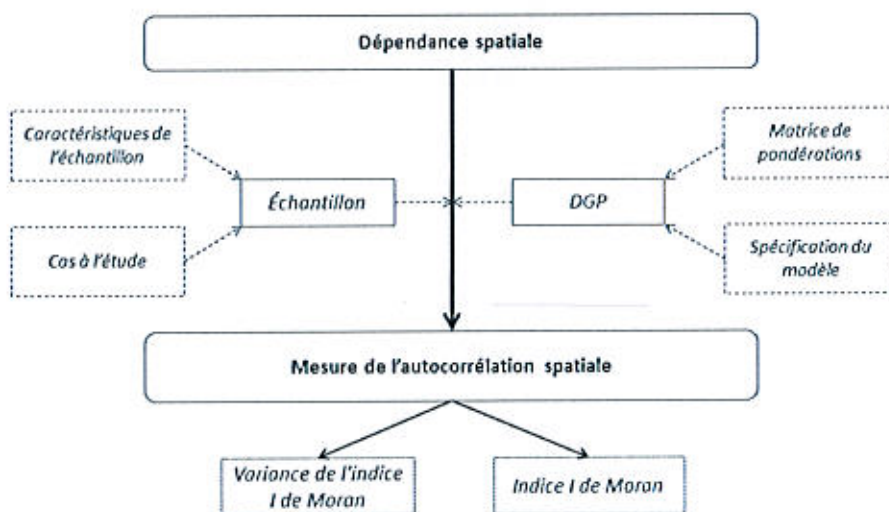
dans le coefficient estimé ainsi que sa variance et donc sur la significativité du test statistique : «Considering the spatial dimension alone when data have an important temporal dimension can lead to the overestimation of spatial dependence in residuals and, consequently, can lead to overcorrection using the appropriate spatial econometric approaches» (Dubé et Legros, 2013a: 8).

Les conclusions des études précédentes ne reposent que sur un nombre limité d'échantillons empiriques¹³. Afin de dépasser cette limite, une première méta-analyse est réalisée par Devaux et Dubé (à venir). Les conclusions mettent également en évidence l'impact d'une spécification strictement spatiale sur l'indice *I* de Moran ainsi que l'influence de l'étendue temporelle rappelant le problème de sur-connectivité déjà mentionné par Smith (2009): «presence of strongly connected spatial weight matrices can introduce serious biases into both the estimation and testing of spatial autocorrelation» (idem, p.25). En revanche, même si cette première méta-analyse contrôle l'effet d'un bon nombre de facteurs (type de matrice de pondérations, forme fonctionnelle, DGP, nombre de variables incluses, etc.) elle présente également certaines limites: i) elle se limite aux approches les plus communément utilisées dans la littérature; ii) elle ne peut considérer que les études rapportant les statistiques spatiales convoitées; iii) un certain nombre d'études recensées ont été rejetées de l'analyse en raison d'une description méthodologique ambiguë et iv) l'harmonisation des données collectées peut être difficile et causer une perte d'informations. L'article propose une nouvelle procédure d'échantillonnage et d'analyse visant à outrepasser ces limites.

2.3 Hypothèses de recherche

On note que l'autocorrélation spatiale origine principalement de deux sources. La première renvoie aux résultats de la modélisation explicite de l'espace et la seconde à une mauvaise spécification économétrique (Anselin and Griffith, 1988; Le Gallo, 2002). Smith (2009) note également que certaines caractéristiques de l'échantillon peuvent également influencer la mesure de la dépendance spatiale. Ainsi comme l'illustre la figure 2, le processus générateur de données (DGP) et l'échantillon seraient deux éléments importants.

Figure 2: Modèle logique de l'analyse



¹³En fait, seuls (Dubé and Legros, 2014b) appliquent cette méthode sur plusieurs (3) échantillons.

Le DGP est défini par la spécification du modèle économétrique et de la matrice de pondérations spatiales/spatiotemporelles. L'échantillon est quant à lui défini par les dynamiques spatiotemporelles comprises dans le cas à l'étude ainsi que par les caractéristiques de la base de données. Les hypothèses de recherche sont les suivantes:

- En supposant la réalisation simultanée des observations et surestimant les relations spatiales, la présence d'autocorrélation spatiale est également surestimée;
- La variance, σ_I^2 , de l'indice de Moran est minimisée lorsque la dimension temporelle est omise;
- Les caractéristiques de l'échantillon affectent la mesure de l'autocorrélation spatiale. L'étendue temporelle, la distance moyenne entre les observation ou leur nombre pourrait influencer les relations spatiales. Cette influence pourrait être accrue lorsqu'aucune contrainte temporelle n'est appliquée aux relations spatiales.

3 Méthodologie

3.1 Mesure de l'autocorrélation spatiale

Parmi les différents tests de détection de la dépendance spatiale, l'indice I de Moran (Moran, 1948, 1950) est celui qui est le plus fréquemment utilisé (Griffith, 2012)¹⁴. La statistique exprime le rapport entre la covariance pondérée pour une variable mesurée à des unités voisines et la variance totale observée (équation 4). Cet indice permet de mesurer la structure spatiale des données et de déterminer si celles-ci sont aléatoirement distribuées dans l'espace (Haining, 2009; Le Gallo, 2002). Il exprime donc l'importance de la différence des valeurs entre toutes les paires d'observations contiguës (Calloz et Collet, 2011) ou proches.

$$I = \frac{N}{S_0} \left(\frac{\bar{\varepsilon}' W \bar{\varepsilon}}{\bar{\varepsilon}' \bar{\varepsilon}} \right) \quad (4)$$

où, N représente le nombre d'observations; S_0 la somme des éléments de la matrice de pondérations ($S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij}$); ε la variable de résidus du modèle de régression OLS; $W\varepsilon$ la variable spatialement décalée basée sur une matrice de pondérations spécifiée de manière exogène (W). Avec une standardisation en ligne de la matrice W la somme de chaque ligne est égale à 1 et $S_0 = N$. L'équation 4 devient alors:

$$I = \frac{\bar{\varepsilon}' W \bar{\varepsilon}}{\bar{\varepsilon}' \bar{\varepsilon}} \quad (5)$$

Ainsi formulé, la statistique de Moran présente de forte similarité avec la statistique de Durbin-Watson (1950) appliquée aux séries chronologiques et où la matrice W remplace la matrice A (Anselin and Rey, 1991; Durbin and Watson, 1950). La standardisation en ligne de la matrice W permet également de simplifier l'interprétation de l'indice d'autocorrélation calculé. Selon le principe d'un coefficient de corrélation l'indice I est alors 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 : pas d'autocorrélation) est basé sur une statistique centrée réduite et normalement distribuée $N(0,1)$ (équation 6):

¹⁴Le Gallo (2002) fait remarquer que si l'autocorrélation spatiale peut résulter d'une spécification économétrique erronée, alors le I de Moran peut également être utilisé comme un indicateur général de spécification

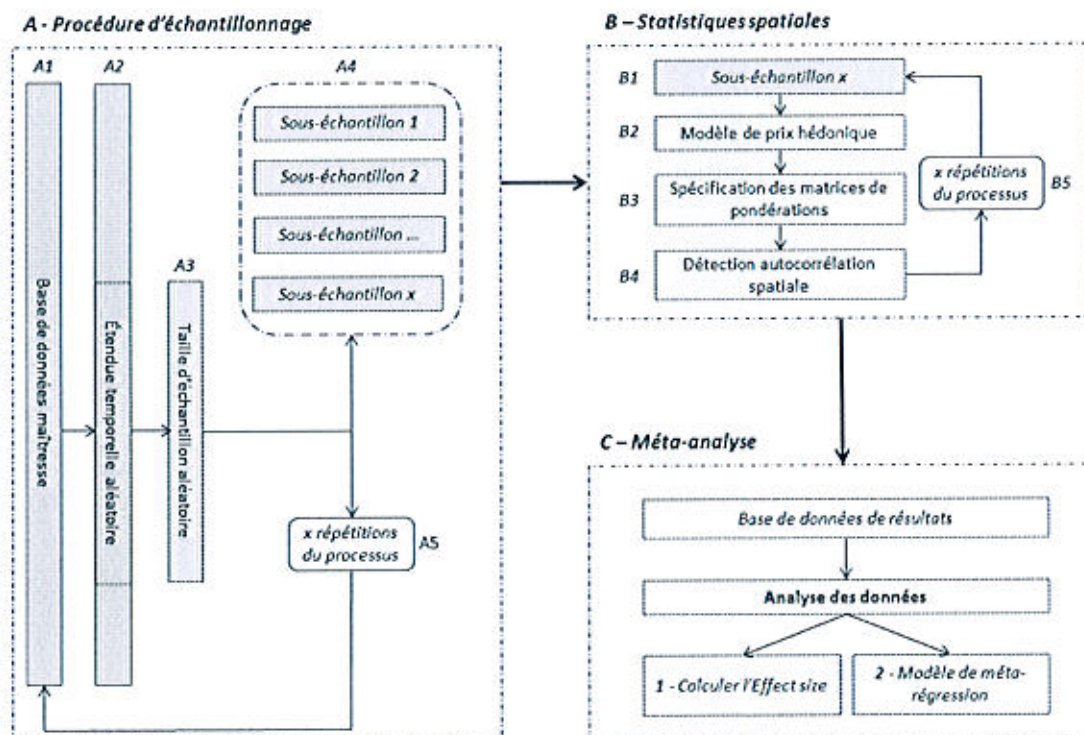
$$z = \frac{I - E(I)}{\sqrt{\sigma_I^2}} \tag{6}$$

où, I représente l'indice de Moran, $E(I)$ son espérance et σ_I^2 sa variance¹⁵.

3.2 Données et stratégie d'analyse

D'ordinaire, le *méta-analyste* n'a pas de contrôle sur la manière dont chaque étude incluse dans son échantillon est réalisée (Pigott, 2012). La méta-analyse de Devaux et Dubé (*à venir*) a mis ce problème en évidence. L'approche préconisée dans cette étude s'affranchit de ces limites et garantit une forme d'homogénéité dans la démarche d'analyse pour chaque échantillons considérés permettant ainsi d'isoler l'effet de certains paramètres qui eux varieront. La démarche d'échantillonnage et d'analyse se décompose en trois étapes: A) Une procédure d'échantillonnage originale est élaborée à partir d'une base de données maîtresse afin de créer une série de sous-échantillons aux caractéristiques variées; B) Les statistiques spatiales nécessaires sont calculées pour chacun des sous-échantillons créés et une nouvelle base de données regroupant les résultats est constituée; C) L'analyse des données est réalisée afin d'isoler et expliquer la variation de la dépendance spatiale mesurée entre les différents sous-échantillons. La figure 3 résume ces étapes:

Figure 3: Procédure d'échantillonnage et d'analyse



¹⁵voir Dubé and Legros (2014a) pour plus de détails sur la calcul de la variance et de l'espérance pour la statistique de Moran

3.2.1 Données utilisées

Les données utilisées regroupent toutes les transactions résidentielles recensées dans la ville de Paris entre 1990 et 2003 (N=135 525) provenant de la "Base d'Informations Économiques Notariales" (BIEN). Pour chaque transaction enregistrée, une adresse permet de géolocaliser le bien vendu et d'obtenir ses coordonnées géographiques précises utiles pour la construction des matrices de pondérations spatiales. Des informations concernant les caractéristiques des logements telles que le type d'appartement, la surface habitable, la date de construction, le nombre de pièce, le nombre de salles de bain, le nombre de garages et l'étage du logement sont disponibles. La démarche méthodologique repose sur la création de sous-échantillon d'étendue temporelle et de taille variables. Pour les fins de l'étude seuls les arrondissements centraux (1 à 11) de Paris sont retenus. En étendant l'aire géographique à toute la ville de Paris, les échantillons créés moins denses sont susceptibles de ne pas entretenir de relations spatiales. Au final, 43 077 observations sont conservées pour la procédure d'échantillonnage. Les statistiques descriptives sont présentées au tableau 1 et la figure 4.

Table 1: Statistiques descriptives: Base de données de départ

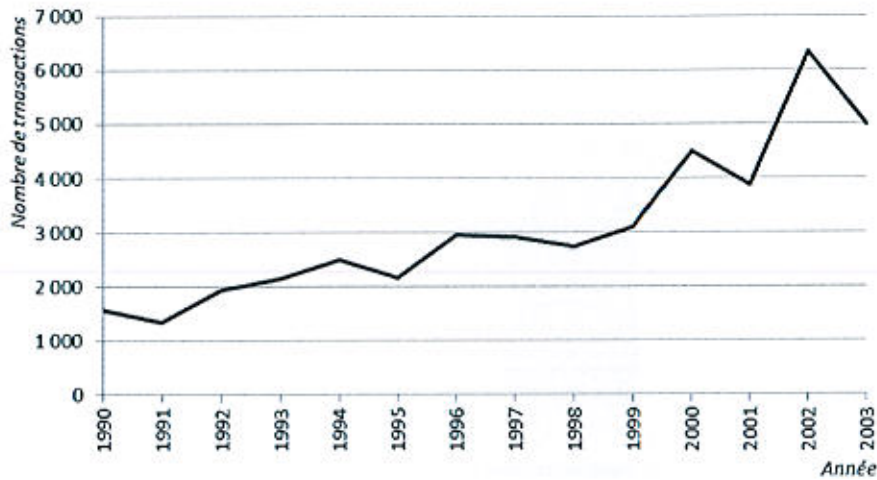
Variable	N	Moyenne	Var.	Type
prix	43077	137041	86860,27	continue
surface	43077	46,0310	25,5669	continue
Nombre de salles de bain	43077	0,5158	0,5360	continue
Ascenseur	43077	0,3487	0,4766	binaire
Chauffage	43077	0,0407	0,1976	binaire
Terrasse	43077	0,0178	0,1323	binaire
Premier étage	43077	0,0544	0,2267	binaire
Deuxième étage	43077	0,1628	0,3692	binaire
Troisième étage	43077	0,1735	0,3787	binaire
Quatrième étage	43077	0,1779	0,3824	binaire
Cinquième étage	43077	0,1703	0,3759	binaire
Sixième étage	43077	0,2611	0,4392	binaire
Construction avant 1850	43077	0,1504	0,3575	binaire
Construction entre 1850 et 1913	43077	0,5193	0,4996	binaire
construction entre 1914 et 1947	43077	0,1110	0,3142	binaire
Construction entre 1948 et 1969	43077	0,0851	0,2791	binaire
Construction entre 1970 et 1980	43077	0,0898	0,2859	binaire
Construction entre 1981 et 1991	43077	0,0218	0,1460	binaire
Construction après 1991	43077	0,0225	0,1484	binaire

3.2.2 Stratégie d'échantillonnage et création de la base de données

La procédure d'échantillonnage s'inspire d'une approche bootstrap. Un programme (Stata) est créé afin de définir les sous-échantillons aux caractéristiques souhaitées (A1 dans la figure 3). Cette approche a l'avantage de produire un nombre souhaité de sous-échantillons tirés d'un contexte similaire mais la portée des résultats devraient être limitée à celui-ci. Une étendue temporelle aléatoire, T_i entre 1 et 60 mois, est déterminée au sein de laquelle un nombre d'observations aléatoire, S entre 1000 et 7000 transactions, sont extraites pour former un sous-échantillon, r (A2 et A3, figure 3). Chaque sous-échantillon (x) créé est stocké (A4) et la procédure est répétée 40 fois (A5).

Un second programme est créé et destiné à calculer les statistiques spatiales souhaitées (I de

Figure 4: Nombre de transactions par an



Moran) pour chacun des sous-échantillons. Les sous-échantillons stockés à l'étape A4 sont isolés et un modèle de prix hédoniques (MPH) pour chacun des échantillons est estimé (B1 et B2):

$$\log(y_{jtx}) = \alpha_{x0} + \sum_{k=1}^K X_{kjtx} \beta_{kjtx} + \varepsilon_{jtx} \quad (7)$$

où y_{jt} représente le prix de transaction réel du bien j à la période t pour le sous-échantillon x . Les valeurs immobilières sont soumises aux fluctuations du marché. Des effets fixes temporels sont généralement inclus afin de capturer ces évolutions et de créer des indices de prix retraçant l'évolution générale du marché par rapport à une période donnée (Wooldridge, 2010). Cependant, comme l'étendue temporelle varie d'un échantillon à l'autre le nombre de variables de contrôle temporelle est également appelé à varier d'un échantillon à l'autre et d'affecter les coefficients estimés. L'utilisation prix réel permet de soustraire l'effet de l'évolution temporelle du marché au prix de vente enregistré. La matrice X représente un ensemble de k variables définissant les caractéristiques intrinsèques du logement ainsi que sa localisation tandis que ε représente le terme d'erreur du modèle de régression. Finalement, α et β_k représentent respectivement la constante et les k coefficients estimés tandis que ε est le terme d'erreur du modèle. La présence de dépendance spatiale n'étant pas contrôlée, ε peut ne pas respecter l'hypothèse d'indépendance.

La présence d'autocorrélation spatiale est calculée à partir des résidus, ε_{jk} , du MPH qui sont intégrés à l'équation 4. Deux catégories de matrices de pondérations sont spécifiées afin de formaliser les relations de dépendance entre les observations: une matrice spatiale (S) et une matrice spatiotemporelle (W). La matrice de pondérations S postule une réalisation simultanée entre toutes les observations. Les relations entre les observations collectées dans l'espace étant non-linéaires (McMillen, 2010), deux types de fonctions décroissantes de la distance¹⁶ sont utilisées afin d'octroyer d'une plus grande pondération aux observations les plus proches. Une première pondère les relations spatiales, S_{inv} , avec une fonction de distance inverse (equation 8) tandis que la seconde, S_{exp} , utilise une fonction exponentielle négative (equation 9). Les deux matrices spatiales sont tronquées à la distance moyenne (\bar{d}).

¹⁶La distance euclidienne entre deux observations i et j (d_{ij}) est déterminée par: $d_{ij} = \sqrt{(lat_i - lat_j)^2 + (long_i - long_j)^2}$.

$$s_{ij}^1 = \begin{cases} d_{ij}^{-\alpha} & \text{si } d_{ij} \leq \bar{d} \quad \forall i \neq j \\ 0 & \text{sinon} \end{cases} \quad (8)$$

$$s_{ij}^2 = \begin{cases} e^{-d_{ij}} & \text{si } d_{ij} \leq \bar{d} \quad \forall i \neq j \\ 0 & \text{sinon} \end{cases} \quad (9)$$

La matrice spatiotemporelle se décline en deux catégories (W_1 et W_2) et sont spécifiées tel que proposé à l'équation 3. Chacune d'elles mise sur une spécification différente de la matrice T destinée à restreindre les relations spatiales à une période donnée. La distance temporelle entre les observations est définie par: $v_{ij} = (v_i - v_j)$ ou v représente le moment de la vente. En pratique, les premières observations collectées ne disposent d'aucune influence de transactions antérieures. Ainsi, l'influence des transactions futures proches est également considérée. Par ailleurs, en intégrant l'effet spatial de ventes futures rapprochées, il est possible de tenir compte d'un effet d'anticipation possible et d'un certain délais d'enregistrement des transactions. La matrice W_1 , contraint les relations spatiales aux ventes enregistrées 30 jours avant et 10 jours après une transactions donnée (équation 10). La matrice W_2 élargit quant à elle les relations aux transactions ayant eu lieu 90 jours plus tôt (équation 11). Au-delà de ces seuils temporels prédéfinis, la pondérations des relations spatiales (d_{ij}) est nulle.

$$t_{ij}^1 = \begin{cases} 1 & \text{si } -10 \text{ jours} < v_{ij} < 30 \text{ jours} \quad \forall i \neq j \\ 1 & \text{si } v_i = v_j \quad \forall i \neq j \\ 0 & \text{sinon} \end{cases} \quad (10)$$

$$t_{ij}^2 = \begin{cases} 1 & \text{si } -10 \text{ jours} < v_{ij} < 90 \text{ jours} \quad \forall i \neq j \\ 1 & \text{si } v_i = v_j \quad \forall i \neq j \\ 0 & \text{sinon} \end{cases} \quad (11)$$

En résumé, 6 matrices de pondérations sont créées (tableau 2).

Table 2: Types de matrices de pondérations utilisées

	Matrice Spatiale	Matrices spatiotemporelles	
<i>Restrictions temporelles (en jours)</i>	Aucune	-30; +10	-90; 0
<i>Distance inverse</i>	S_{inv}	W_{1inv}	W_{2inv}
<i>Distance Exp. Négative</i>	S_{exp}	W_{1exp}	W_{2exp}

Finalement, le I de Moran est calculé à partir de chacune des matrices de pondérations créées (B4) et la procédure est répétée pour chacun des sous-échantillons (B5). La fonction de distance inverse est appliquée à la moitié des échantillons créés et la fonction de distance négative exponentielle à l'autre moitié. Les résultats sont stockés dans une nouvelle base de données comprenant autant d'observations que de sous-échantillons créés à l'étape A. Pour chaque observation, sont stockées les statistiques de Moran calculées, des informations concernant l'échantillon de départ (taille, étendue temporelle, distance moyenne entre les observations) et le type de fonction de distance utilisée (tableau 4).

Table 3: Statistiques descriptives

Variable	Définition	Type
<i>Range</i>	Étendue temporelle du sous-échantillon (en mois)	Continue
<i>Ranged</i>	Étendue temporelle du sous-échantillon (en jours)	Continue
<i>lnRanged</i>	log de l'étendue temporelle du sous-échantillon (en jours)	Continue
<i>Dmean</i>	Distance moyenne entre les observations. Au-delà de cette distance, les relations spatiales sont considérées comme nulle (en mètres)	Continue
<i>SampSize</i>	Nombre d'observations dans le sous-échantillon	Continue
<i>DistFunc</i>	Type de fonction de la distance utilisée pour spécifier la matrice de pondérations spatiales (inverse:1; nég. exp.: 0)	Binaire
$R^2 - ajust$	R^2 -ajusté du modèle de prix hédoniques (MPH)	Continue
R^2	R^2 du modèle de prix hédoniques (MPH)	Continue
<i>Variance I</i>	Variance de l'indice I de Moran estimé pour chaque sous-échantillon	Continue
I^*	Indice I de Moran estimé selon le type de matrice de pondérations spécifié (*)	Continue
<i>StdErr*</i>	Écart-type de l'indice I de Moran selon le type de matrice de pondérations spécifié (*)	Continue

Table 4: Statistiques descriptives

	N	Mean	SE	Min	Max
<i>I de Moran, Matrice S</i>	40	0,0039	0,0032	-0,0036	0,0101
<i>I de Moran, Matrice W₁</i>	40	0,0242	0,0335	-0,0029	0,1248
<i>I de Moran, Matrice W₂</i>	40	0,0197	0,0253	-0,0029	0,0941
<i>Écart-type I de Moran, Matrice S</i>	40	0,0011	0,0006	0,0004	0,0023
<i>Écart-type I de Moran, Matrice W₁</i>	40	0,0013	0,0007	0,0005	0,0035
<i>Écart-type I de Moran, Matrice W₂</i>	40	0,0013	0,0007	0,0005	0,0033
R^2	40	0,8301	0,0269	0,7846	0,8680
R^2 -ajust.	40	0,1714	0,0260	0,1351	0,2132
<i>Taille de l'échantillon (S_x)</i>	40	3199	1520	1013	6895
<i>Étendue temporelle (en mois)</i>	40	31	15	4	58
<i>Étendue temporelle (en jours)</i>	40	997	477	152	1827
<i>Distance moyenne entre les observations</i>	40	2,5733	0,0461	2,4960	2,6590
<i>Type de fonction de distance (1: $d_{ij}^{-\alpha}$; 0: $e^{-d_{ij}}$)</i>	40	0,5	0,506	0	1

3.2.3 Effect size et modèle de méta-régression

À partir d'une problématique donnée, les méta-analyses visent généralement à estimer un effet moyen calculé pour l'ensemble des études recensées et à expliquer ses fluctuations d'une étude à l'autre. Les sous-échantillons créés remplacent les études empiriques collectées dans le cas d'une revue de la littérature *traditionnelle*. Le déroulement d'une méta-analyse se compose généralement de deux étapes: i) estimer un effet moyen (*effect size*, *ES*) pour l'ensemble des études utilisées; ii) modéliser cet effet.

L'*Effect Size*, *ES*, correspond à la moyenne de l'effet mesuré pour une étude donnée et pondérée par l'inverse de sa variance (variance intra-échantillon) de sorte à octroyer plus d'importance aux estimations les plus précises. Tel que présenté par Borenstein et al. (2009), l'effet moyen pondéré est défini par l'équation 14:

$$ES = \frac{\sum_{x=1}^k P_x I_x}{\sum_{x=1}^k P_x} \quad (12)$$

où I_x représente l'effet considéré (Indice I de Moran) pour sous-échantillon x et P_x est le poids associé calculé par l'inverse de la variance de l'effet mesuré, $1/v_{I_x}$. L'effet mesuré correspondant à la mesure d'autocorrélation spatiale qui peut en partie varier d'un échantillon à l'autre de manière aléatoire (structure spatiale différente, des caractéristiques des bien vendus, etc). La statistique Q permet de tester la présence d'hétérogénéité dans la variation des indices de Moran mesurés (H_0 = effets homogènes)¹⁷ d'un sous-échantillon à l'autre. Elle est interprétée comme une comparaison de la variance inter-études et intra-études de l'effet (Hedges and Vevea, 1998):

$$Q = \sum_{x=1}^k P_x I_x^2 - \frac{(\sum_{x=1}^k P_x I_x)^2}{\sum_{x=1}^k P_x} \quad (13)$$

Dans le cas où l'effet mesuré est statistiquement non homogène entre les sous-échantillons la pondération de *ES* est alors calculée par l'inverse de la somme de la variance intra-échantillon, v_{I_x} , et inter-échantillon, τ^2 : $P_x^* = \frac{1}{v_{I_x} + \tau^2}$. L'ajout du paramètre τ^2 permet de capter ces variations de mesures qui ne sont pas uniquement liées à la stratégie d'échantillonnage¹⁸. Le nouvel effet moyen pondéré, ES^* , sera alors obtenu par:

$$ES^* = \frac{\sum_{x=1}^k P_x^* I_x}{\sum_{x=1}^k P_x^*} \quad (14)$$

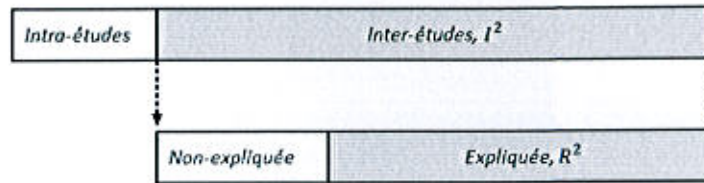
La réalisation d'une méta-analyse ne se limite pas à la mesure d'un effet moyen (*ES*). Elle vise également à identifier et expliquer en quoi, d'une étude à l'autre, la variation d'une statistique peut être associée à ses caractéristiques (Harbord and Higgins, 2008). En d'autres termes, l'estimation d'un modèle de méta-régression vise à distinguer et expliquer le pourcentage de variation inter-études, notée I^2 , compris dans la variation totale de l'effet mesuré (Higgins et al., 2002). Borenstein et al. (2009) précisent que cette statistique devrait être interprétée comme une mesure d'inconsistance (*inconsistency*) dans les résultats des études ciblées par une méta-analyse (voir figure 5¹⁹).

¹⁷Dans le cas où les effets mesurés sont identiques (H_0 est vraie), alors Q suit une distribution Chi-2 à $k - 1$ degrés de liberté. Donc, si la statistique Q dépasse la valeur critique à $k - 1$ degrés de liberté, on rejette l'hypothèse H_0 .

¹⁸Plusieurs auteurs discutent de l'estimation du paramètre τ^2 . Pour plus de détails, le lecteur intéressé pourra notamment se référer à Pigott (2012); Hedges and Vevea (1998).

¹⁹Figure tirée de (Borenstein et al., 2009, p.202).

Figure 5: Variation inter-études et intra-études



Dans le cas présent, la le I^2 exprime donc les variations de la mesure de la dépendance spatiales qui ne sont pas attribuables aux processus spatiotemporels partagés par tous les sous-échantillons et propres au marché immobilier parisien pour la zone ciblée. Cette variation peut être expliquée par les caractéristiques de l'échantillon ou la stratégie de modélisation. Le modèle de prix hédoniques étant identique pour chaque sous-échantillon une part de cette variation est attribuable à la matrice de pondérations spécifiée. En revanche, une autre part du I^2 peut être d'origine aléatoire.

-> Trouver un moyen de distinguer le I de Moran du I^2

Le modèle de méta-régression a effet aléatoire permet de capturer la présence d'hétérogénéité dans les résidus qui n'est pas expliquée par un ensemble de covariables et assume que le véritable effet suit une distribution normale. Tel que présenté par Harbord and Higgins (2008), le modèle prendra la forme suivante:

$$I_{xw} = M_{xwz} \phi_{xwz} + u_{xw} + \epsilon_{xw} \quad (15)$$

où, I_{xw} représente l'effet mesuré (I de Moran) pour chaque sous-échantillon x selon le type de matrice de pondérations utilisé w ; M est un ensemble de z covariables; u est le terme d'effet aléatoire de moyenne nulle et de variance τ^2 et ϵ est le terme d'erreur indépendant et identiquement distribué. Au final, les coefficients estimés expliquent la part de variation inter-échantillons une fois que la variation causée par l'hétérogénéité est captée.

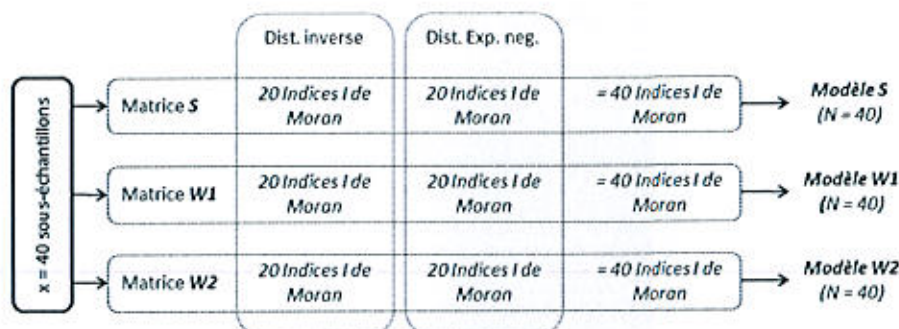
4 Analyse et résultats

4.1 Mesure de la taille de l'effet

L'estimation d'un effet moyen (*effect size*) vise à établir si une différence significative est notable entre différentes matrices de pondérations spatiales. Cette différence peut s'exprimer en termes de valeur ou de précision de l'indice de Moran (variance, σ_I^2). L'intuition de départ suggère que sans restriction temporelle les relations spatiales sont faussées occasionnant une surévaluation de l'autocorrélation spatiale et une tendance à accroître la précision de cette mesure affectant les tests d'hypothèses. L'effet moyen pondéré est estimé à partir des résidus du modèle de prix hédoniques estimés pour chaque échantillon et avec deux fonctions de distances continues distinctes. Au total, 40 sous-échantillons permettent de calculer 120 indices I de Moran et leur variance à partir des matrices de pondérations S , W_1 et W_2 et des fonctions de distance inverse et exponentielle négative (figure 6).

Le tableau 5 présente les effets moyens pondérés (ES^*) ainsi que les intervalles de confiance inférieurs (LL) et supérieurs (UL) associés. Les indices I de Moran moyens rapportés mesurent une

Figure 6: Calcul des I de Moran selon le type de matrice de pondérations



présence significative d'autocorrélation spatiale (colonne 3 du tableau 5). La présence d'hétérogénéité affecte significativement leur variation et justifie l'utilisation du modèle présenté à l'équation 14. Contrairement à l'intuition de départ, l'indice *I* de Moran se révèle en moyenne plus faible lorsqu'une matrice de pondérations strictement spatiale est utilisée. En fait, plus les relations spatiales autorisées intègrent une fenêtre temporelle de courte durée et plus la mesure de l'autocorrélation se révèle élevée. La figure 7 illustre les résultats.

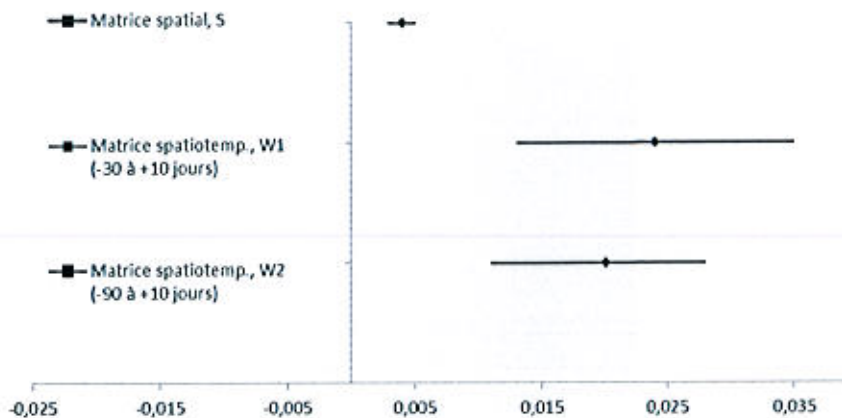
Table 5: Effet moyen pondéré de l'Indice de Moran selon le type de matrice de pondérations

		Effect Size	sig. ES/Heterog.
Matrice spatiale <i>sans restriction temporelle</i>	ES_S	0,004	*** / ***
	LL_S	0,003	
	UL_S	0,005	
Matrice spatiotemporelle 1 <i>restrict. -30 et +10 jours</i>	ES_{W1}	0,024	*** / ***
	LL_{W1}	0,013	
	UL_{W1}	0,035	
Matrice spatiotemporelle 2 <i>restrict. -90 jours</i>	ES_{W2}	0,020	*** / ***
	LL_{W2}	0,011	
	UL_{W2}	0,028	

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

L'effet moyen pondéré estimé pour chaque matrice de pondérations *y* est représenté par un carré tandis que les bornes gauche et droite de chaque ligne représentent respectivement l'intervalle de confiance inférieur et supérieur. Si les intervalles de confiance (lignes) se superposent pour deux effets moyens distincts, alors il n'existe pas de différence significative entre ceux-ci. On note à cet effet que les deux types de matrices de pondérations spatiotemporelles (W_1 et W_2) présentent des mesures significativement différentes de l'effet mesuré lorsqu'une matrice de pondérations strictement spatiale est spécifiée. En revanche, cette observation ne s'applique pas aux matrices spatiotemporelles entre elles. L'indice *I* de Moran tend à diminuer à mesure que la fenêtre temporelle s'accroît mais les effets moyens ne sont pas significativement différents. Conformément à l'intuition de départ, l'utilisation d'une matrice de pondérations strictement spatiale présente des intervalles de confiance en moyenne plus faibles. Cette observation rejoint les résultats de Dubé et Legros (2013a) qui rappellent qu'une variance plus faible pourrait affecter les tests statistiques en orientant le chercheur vers un rejet abusif de l'hypothèse nulle (H_0 : pas d'autocorrélation spatiale).

Figure 7: Effet moyen pondéré du I de Moran selon le type de matrice de pondérations



4.2 Modèle de méta-régression

Un modèle de régression à effet aléatoire est estimé afin d'expliquer la variation de l'indice I de Moran entre les différents échantillons tout en captant la part de variation affectant la distribution homogène des résidus. Trois modèles sont estimés et comparés selon le type de matrice de pondérations utilisé pour estimer l'indice I de Moran. Les caractéristiques de chaque sous-échantillon (étendue temporelle, distance moyenne, nombre de transactions) ainsi que le type de fonction de distance font office de variables indépendantes. Le tableau 6 présente les résultats d'estimations.

Table 6: Résultats d'estimations

Variables	Matrice S		Matrice W_1		Matrice W_2	
	Coef.	sig.	Coef.	sig.	Coef.	sig.
Sample time range (in days)	-5,34E-07		1,08E-04	***	7,64E-05	***
log of Sample time range	0,0008		-0,0594	**	-0,0409	**
Sample's mean dist. between obs. (m)	-0,0265	**	0,2063		0,1304	
Sample's number of observations	6,66E-07	**	5,55E-06	*	4,60E-06	*
Distance function (inverse:1; neg.exp.:0)	0,0039	***	0,0070		0,0064	
Adjusted R^2	-0,0546	**	0,4658	*	0,3227	*
Constant	0,0725	**	-0,3139		-0,1889	
N	40,00		40,00		40,00	
τ^2	1,70E-06		0,000397		0,000232	
I^2	72,02		99,77		99,63	
Adj - R^2	77,71		64,59		63,64	

Notes: * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$;
 τ^2 : estimation de la variance inter-échantillon;
 I^2 : % de variation due à l'hétérogénéité;
 Adj - R^2 : % de la variance inter-échantillon expliquée.

Tel que présenté à la figure 5, la statistique I^2 représente la part de variation inter-échantillons (observations) causée par l'hétérogénéité. Une valeur de I^2 proche de 0 suggère que la distribution de la variable expliquée est d'origine aléatoire. Inversement, en s'éloignant de 0 la statistique suggère qu'une part de la variance des «inconsistances» entre les sous-échantillons est explicable par un ensemble de covariables (Borenstein et al., 2009, p.122). Chacun des modèles présentés présente une

statistique I^2 élevée (variant entre 72,02% et 99,7%. Le R^2 -Ajusté présente la part de variation du I^2 expliquée par le modèle tandis que τ^2 rapporte la variance résiduelle associée au paramètre aléatoire du modèle (Harbord and Higgins, 2008, p.77). Le I^2 élevé suppose que la variation de mesure de l'autocorrélation n'est pas causée par des changements dans la structure des processus spatiaux d'un échantillon à l'autre. Cela s'explique par l'origine commune des sous-échantillons créés provenant de la même base de données (transactions parisiennes). La statistique sera en revanche plus faible dans le cas du premier modèle (matrice S). La détection de l'autocorrélation a été décrite comme une tâche complexe en raison des facteurs propres aux données ou à la spécification du modèle de régression susceptibles d'influencer sa mesure. On note en premier lieu un effet significatif négatif (-0,0265) de la distance moyenne entre les observations sur la statistique d'intérêt lorsque la dimension temporelle est omise (modèle Matrice S). À l'inverse, la distance moyenne entre les observations n'exerce plus d'effet lorsque des restrictions temporelles sont considérées (modèles W_1 et W_2). Par ailleurs, la forme fonctionnelle de la distance présente des résultats similaires. Aucune distinction n'est faite entre une matrice basée sur l'inverse ou l'exponentiel négatif de la distance lorsque la matrice de pondérations est temporellement contrainte. En revanche, dans un contexte strictement spatial, la fonction inverse octroie un poids supérieur aux observations les plus proches et augmente significativement la mesure de l'autocorrélation spatiale (0,0039).

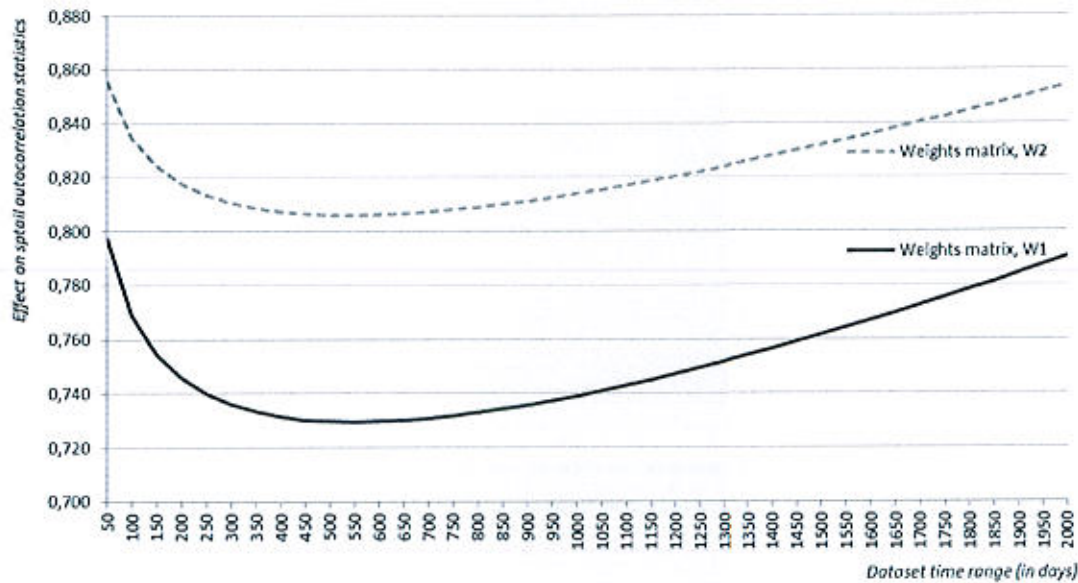
La statistique du R^2 ajusté du modèle de prix hédonique estimé pour chaque régression est intégrée comme variable dépendante. Un pouvoir explicatif élevé suppose une variation résiduelle moindre. Un R^2 ajusté supérieur a tendance à diminuer la mesure de la dépendance spatiale dans le cas d'une matrice spatiale (-0,0546). En revanche, l'indice I de Moran tend à être fortement augmenté pour les deux autres spécifications (modèles W_1 et W_2). Les résidus du modèle MCO présentent donc une autocorrélation plus grande lorsqu'une matrice spatiotemporelle est spécifiée et suggère que les mesures spatialement et temporellement proches ont plus tendance à se ressembler. Il est aussi possible de supposer qu'en assumant une réalisation simultanée des mesures, des relations spatiales sont postulées entre des transactions spatialement proches mais dont les valeurs se distinguent en raison de leur évolution dans le temps.

L'étendue temporelle de la base de données influence significativement la mesure de l'autocorrélation spatiale lorsqu'une matrice spatio-temporelle est spécifiée. Afin de vérifier la présence d'une possible relation non-linéaire, la transformation logarithmique de la variable d'étendue temporelle est également intégré aux modèles. L'effet non-linéaire est vérifié par un coefficient significatif et illustré à la figure 8. La dépendance spatiale mesurée est décroissante jusqu'à une étendue temporelle de 549 et 535 jours pour les matrice W_1 et W_2 . Dans le cadre d'une spécification uniquement spatiale, l'effet est négatif mais non significatif.

Finalement, le nombre d'observations d'un échantillon influence positivement la détection de l'autocorrélation spatiale pour les trois modèles considérés. Un nombre croissant d'observations favoriserait les connectivités possibles entre les observations et favoriserait la mise en relation de mesures spatialement similaires.

5 Discussion

La méthode d'échantillonnage a été développée pour répondre à certaines limites posées par une méta-analyse *conventionnelle* basée sur une revue de la littérature. L'approche proposée a permis la construction de sous-échantillons et l'obtention de statistiques spatiales plus homogènes à par-

Figure 8: Effet non-linéaire de l'étendue temporelle sur l'indice I de Moran

tir d'une base de données unique, mais dont les paramètres et conditions d'échantillonnage varient. En effets, les données utilisées intègrent des processus spatiaux empiriques et les méthodes d'estimations demeurent comparables d'un sous-échantillon à l'autre. L'harmonisation des données collectées est simplifiée et limite la perte d'informations. Elle permet aussi de centrer l'analyse sur les éléments d'intérêts sans se restreindre aux approches de modélisation les plus communes. En revanche, la portée des résultats devraient être limitées à la base de données de base (transactions immobilières parisiennes).

Les résultats présentés permettent de distinguer l'effet d'une spécification strictement spatiale de la matrice de pondérations sur la détection de l'autocorrelation spatiale. En moyenne, l'indice I de Moran se révèle plus faible lorsque la temporalité des phénomènes est omise. Il serait alors envisageable que les observations, réparties dans l'espace et le temps, reflètent les fluctuations du marché. En d'autres termes, l'omission de la dimension temporelle permettrait la mesure de valeurs qui, même en étant spatialement proches, auraient tendance à être différentes supposant alors que les effets de l'autocorrélation ne sont pas constants dans le temps. En revanche, la spécification d'une matrice de pondérations spatiotemporelles aurait l'effet inverse. Les relations spatiales restreintes à une fenêtre temporelle donnée auront tendance à avoir une plus grande similarité spatiale des réalisations. Une des hypothèses de départ suggère que l'omission de la dimension temporelle augmente les relations spatiales et tend à surestimer la présence d'autocorrélation. Les résultats montrent que les mesures spatialement proches ont tendance à se ressembler davantage si elles sont également temporellement proches. Ce constat fait également sens dans le contexte du marché immobilier parisien volatile. Les coefficients associés à l'impact de l'étendue temporelle soutiennent cette hypothèse.

La détection de la dépendance spatiale est affectée de manière significative mais négative lorsque les observations sont considérées comme étant simultanées (matrice S) et de manière positive dans le cas contraire (matrices W_1 à W_2). Les restrictions temporelles permettraient donc de restreindre les relations spatiales à des observations qui ont tendance à être semblables. D'ailleurs, l'indice I de Moran diminue lorsque la fenêtre temporelle des relations spatiales augmente. Dans le pro-

longement des travaux de Dubé et Legros (2013b; 2013a), une seconde hypothèse suggère que la précision (variance, σ_f^2) de la mesure de l'autocorrélation est affectée par l'omission de la dimension temporelle. Les résultats associés à l'effet moyen semblent confirmer cette hypothèse et peuvent être expliqués par le nombre important de relations spatiales permises par la matrice de pondérations S .

Le modèle de méta-régression visait à expliquer les variations observées entre les mesures de l'autocorrélation spatiale pour les différents sous-échantillons. Les estimations tendent à souligner l'impact des caractéristiques de l'échantillon lorsque les relations spatiales sont surestimées et suppose que la structure spatiale des observations, en partie responsable des processus générateurs de dépendance spatiale, est relayées en arrière plan. Si tel est le cas, la détection de l'autocorrélation serait faussée par l'importance des connectivités et les méthodes de correction pourraient également être affectées. Une étude de cas possédant un grand nombre d'observations aura tendance à rapporter une plus grande autocorrélation spatiale si une matrice strictement spatiale est spécifiée. Le caractère exogène de la matrice de pondérations spatiales fait débat.

Le choix de la fonction de distance inverse octroie une poids supérieur aux observations proches en comparaison à une matrice basée sur une distance exponentielle négative. Ce résultat renforce d'autant plus le rôle de la connectivité. Comme Smith (2009, p.25) le fait remarquer, la présence d'une dépendance statistique réduit la quantité d'information fournie pour chaque observation individuelle. En conséquence, une forte connectivité spatiale reflète une forte dépendance entre les observations et affectera les estimations et tests d'hypothèses. Par ailleurs, il a été mentionné que le nombre de transactions immobilières est susceptible de varier pour chaque période de temps (les biens ne sont pas vendus à intervalle régulier) mais chaque échantillon est susceptible de présenter des biens vendus à plusieurs reprises. L'information redondante influence la présence d'autocorrélation (Griffith, 2005) et est plus susceptible d'être présente lorsque la dimension temporelle est ignorée. Finalement, le rôle des caractéristiques de l'échantillon peut être souligné par le paramètre τ^2 qui exprime la variance du paramètre aléatoire. Une variance plus grande sera alors synonyme d'une part plus grande de variation de la statistique de Moran non associée aux covariables qui décrivent l'échantillon. Celle-ci se révèle plus faible lorsque la dimension temporelle est ignorée ce qui supporte la présence de biais causés sur la mesure de la dépendance spatiale et causée par l'omission du temps.

6 Conclusion

Le projet de recherche vise à apporter des éléments de réflexion sur la formalisation des processus de dépendance qui structurent les observations. La modélisation des processus spatiotemporels de dépendance pour les micro-données collectées de manière non répétée est un enjeu actuel important. Pourtant, les méthodes de modélisation demeurent variées et certaines de leurs principales limites ont été exposées (problèmes d'agrégation, d'aires modifiables, etc.). Quelques études ponctuelles ont été réalisées afin d'évaluer l'impact de l'omission de la dépendance temporelle dans l'analyse de telles données mais, à notre connaissance, ces travaux se limitent à un nombre restreint d'échantillons.

La méta-analyse peut être comprise comme un temps de pause et de synthèse dans la production massive d'études de cas. Elle permet d'évaluer l'effet des caractéristiques de chaque études. Une approche originale d'échantillonnage a été développée en réponse aux limites mises en évidence par à une première revue de la littérature *classique* (Devaux et Dubé, à venir). La nouvelle démarche

proposée s'inspire de la méta-analyse mais se distingue en se basant sur un échantillon de départ unique. Elle permet donc de tester des approches de modélisation à partir d'un cas précis tout en faisant varier certaines caractéristiques des données utilisées. Il est ainsi possible d'avoir accès à des statistiques peu rapportées dans la littérature et d'évaluation de méthodes de modélisation alternatives dans un contexte particulier. Les sous-échantillons créés ont été utilisés pour estimer des modèles de prix hédoniques et tester la présence d'autocorrélation spatiale sur les résidus comme l'aurait fait une étude de cas isolée.

En estimant un effet moyen (*effect size*) et un modèle de méta-régression, l'analyse des résultats visait à expliquer les variations de mesures de la dépendance spatiale (indice I de Moran) selon: i) le type de matrice de pondérations spécifiée et ii) les caractéristiques des échantillons. Des hypothèses de départ ont été formulées. La première supposait que l'omission de la dimension temporelle (hypothèse de réalisation simultanée des observations) engendre une surestimation des relations spatiales et donc une surestimation de la dépendance spatiale. Les résultats tendent à montrer le contraire puisque la dépendance spatiale tend à être en moyenne significativement moindre lorsqu'une matrice de pondérations strictement spatiale est spécifiée. En postulant des relations entre des observations futures et passées, celles-ci pourraient fluctuer au rythme du marché et se révéler trop différentes pour être corrélées même si elles sont spatialement proches.

La seconde hypothèse supposait que la variance de l'indice I de Moran serait minimisée en cas d'omission de la dimension temporelle. Les résultats vérifient cette seconde proposition. La variance plus faible tend à orienter le chercheur vers le rejet de l'hypothèse nulle d'absence d'autocorrélation spatiale. Finalement, il a été supposé que les caractéristiques de l'échantillon auraient une influence sur l'estimation de la dépendance spatiale. Ce point a été partiellement vérifié. La dépendance spatiale mesurée en utilisant une approche strictement spatiale aurait en effet tendance à être significativement influencée par l'échantillon (étendue temporelle, nombre d'observations, distance moyenne entre les observations). L'approche spatiotemporelle basée misant sur une matrice de pondérations générale W serait donc moins soumise aux problèmes de connectivité et de taille d'échantillon efficace (*effective sample size*) mentionnés par Smith (2009).

En résumé, l'approche d'échantillonnage proposée semble prometteuse pour comparer les méthodes de modélisation. Les conclusions invitent à orienter le chercheur vers une approche spatiotemporelle qui semble capter plus adéquatement la présence de dépendance spatiale et rendre cette mesure moins sensible aux caractéristiques de l'échantillon, du marché ou au type de fonction de distance utilisé. Finalement, ces résultats se limitent au contexte des données de Parisiennes. D'autres bases de données pourraient être incluses dans la construction des sous-échantillons. Des analyses similaires à venir devraient désormais porter sur d'autres statistiques et paramètres (tests de spécification et coefficients autorégressifs).

Bibliographie

- Alberto, I., Beamonte, A., Gargallo, P., Mateo, P. M., and Salvador, M. (2010). Variable selection in star models with neighbourhood effects using genetic algorithms. *Journal of Forecasting*, 29(8):728–750.
- Anselin, L. (1988). *Spatial econometrics: methods and models*, volume 4. Springer.
- Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, 27(3):247–267.
- Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1):3–25.
- Anselin, L. and Florax, R. J. (1995). *Small sample properties of tests for spatial dependence in regression models: Some further results*, pages 21–74. Springer.
- Anselin, L. and Griffith, D. A. (1988). Do spatial effects really matter in regression analysis? *Papers in Regional Science*, 65(1):11–34.
- Anselin, L., Le Gallo, J., and Jayet, H. (2006). *Spatial econometrics and panel data models*. by L. Matyas, and P. Sevestre. Kluwer, Dordrecht.
- Anselin, L. and Rey, S. (1991). Properties of tests for spatial dependence in linear regression models. *Geographical analysis*, 23(2):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(4):377–395.
- Baltagi, B. H., Bratberg, E., and Holmås, T. H. (2005). A panel data study of physicians' labor supply: the case of norway. *Health Economics*, 14(10):1035–1045.
- Baltagi, B. H., Song, S. H., and Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of econometrics*, 117(1):123–150.
- Beamonte, A., Gargallo, P., and Salvador, M. (2008). Bayesian inference in star models using neighbourhood effects. *Statistical Modelling*, 8(3):285–311.
- Beamonte, A., Gargallo, P., and Salvador, M. (2010a). Analysis of housing price by means of star models with neighbourhood effects: A bayesian approach. *Journal of Geographical Systems*, 12(2):227–240.
- Beamonte, A., Gargallo, P., and Salvador, M. (2010b). Robust bayesian inference in star models with neighbourhood effects. *Journal of Statistical Planning and Inference*, 140(11):3047–3057. Spatio-temporel.
- Bhattacharjee, A., Castro, E., and Marques, J. (2012). Spatial interactions in hedonic pricing models: The urban housing market of aveiro, portugal. *Spatial Economic Analysis*, 7(1):133–167.
- Borenstein, M., Hedges, L. V., Higgins, J., and Rothstein, H. R. (2009). *Introduction to Meta-analysis*. John Wiley and sons.
- Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional science and urban economics*, 22(3):453–474.

- Chen, Y. (2012). On the four types of weight functions for spatial contiguity matrix. *Letters in Spatial and Resource Sciences*, 5(2):65–72.
- Cliff, A. D. and Ord, J. K. (1973). *Spatial Autocorrelation*. Pion, London.
- Cressie, N. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. John Wiley and Sons.
- Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of econometrics*, 30(1):109–126.
- Debrezion, G., Pels, E., and Rietveld, P. (2007). The impact of railway stations on residential and commercial property value: A meta-analysis. *Journal of Real Estate Finance and Economics*, 35(2):161–180.
- Dubé, J. and Legros, D. (2013a). Dealing with spatial data pooled over time in statistical models. *Letters in Spatial and Resource Sciences*, 6(1):1–18.
- Dubé, J. and Legros, D. (2013b). A spatio-temporal measure of spatial dependence: An example using real estate data. *Papers in Regional Science*, 92(1):19–30.
- Dubé, J. and Legros, D. (2014a). *Économétrie spatiale appliquée des microdonnées*. Collection Système d'Information Géographique. ISTE editions, London.
- Dubé, J. and Legros, D. (2014b). Spatial econometrics and spatial data pooled over time: Towards an adapted modelling approach. *Journal of Real Estate Literature*, 22(1):101–126.
- Dubé, J. and Legros, D. (2014c). Spatial econometrics and the hedonic pricing model: what about the temporal dimension? *Journal of Property Research*, 31(4):333–359.
- Dubé, J. and Legros, D. (2014d). Spatial econometrics and the hedonic pricing model: what about the temporal dimension? *Journal of Property Research*, 31(4):333–359.
- Dubin, R. A. (1998). Spatial autocorrelation: A primer. *Journal of Housing Economics*, 7(4):304–327.
- Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression. *Biometrika*, 37(3-4):409–428.
- Elhorst, J. P. (2014a). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer.
- Elhorst, J. P. (2014b). *Spatial panel data models*, pages 37–93. Springer.
- Fisher, R. A. (1935). The design of experiments.
- Ghinamo, M. L. (2012). Explaining the variation in the empirical estimates of academic knowledge spillovers. *Journal of Regional Science*, 52(4):606–634.
- Gibbons, S. and Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, 52(2):172–191.
- Griffith, D. A. (2005). Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the Association of American Geographers*, 95(4):740–760.
- Griffith, D. A. (2012). Spatial statistics: A quantitative geographer's perspective. *Spatial Statistics*, 1(0):3–15.

- Haining, R. P. (2009). Spatial autocorrelation and the quantitative revolution. *Geographical Analysis*, 41(4):364–374.
- Harbord, R. M. and Higgins, J. (2008). Meta-regression in stata. *Meta*, 8(4):493–519.
- Heckman, J. J. and Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of econometrics*, 30(1):239–267.
- Hedges, L. V. and Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods*, 3(4):486.
- Higgins, J. P., Thompson, S. G., et al. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11):1539–1558.
- Huang, B., Wu, B., and Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3):383–401.
- Le Gallo, J. (2002). Économétrie spatiale : l'autocorrélation spatiale dans les modèles de régression linéaire. *Economie et Prevision*, 155(4):139–157.
- LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton, FL: Chapman and Hall/CRC.
- LeSage, J. P. and Pace, R. K. (2010). The biggest myth in spatial econometrics. Available at SSRN 1725503.
- Liu, X. (2013). Spatial and temporal dependence in house price prediction. *Journal of Real Estate Finance and Economics*, 47(2):341–369.
- McMillen, D. P. (2010). Issues in spatial data analysis. *Journal of Regional Science*, 50(1):119–141.
- Moffitt, R. (1993). Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics*, 59(1):99–123.
- Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):243–251.
- Moran, P. A. (1950). A test for the serial independence of residuals. *Biometrika*, pages 178–181.
- Nappi-Choulet, I. and Maury, T. P. (2011). A spatial and temporal autoregressive local estimation for the paris housing market. *Journal of Regional Science*, 51(4):732–750.
- Nappi-Choulet, I. Pr., M. T. P. (2009). A spatiotemporal autoregressive price index for the paris office property market. *Real Estate Economics*, 37(2):305–340.
- Nelson, J. P. (2004). Meta-analysis of airport noise and hedonic property values: Problems and prospects. *Journal of Transport Economics and Policy*, 38(1):1–27.
- Openshaw, S. (1983). *The modifiable areal unit problem*, volume 38. Geo books Norwich.
- Pace, R. Kelley Lesage, J. P. (2004). Spatial statistics and real estate. *Journal of Real Estate Finance and Economics*, 29(2):147–148.

- Pace, R. K., Barry, R., Clapp, J. M., and Rodriguez, M. (1998). Spatiotemporal autoregressive models of neighborhood effects. *Journal of Real Estate Finance and Economics*, 17(1):15–33.
- Pace, R. K., Barry, R., Gilley, O. W., and Sirmans, C. F. (2000). A method for spatial–temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, 16(2):229–246.
- Pigott, T. (2012). *Advances in meta-analysis*. Springer Science and Business Media.
- Schipper, Y., Nijkamp, P., and Rietveld, P. (2002). *Thirty years of aircraft noise value studies: a meta-analysis*, chapter 12, pages 235–245. Edward Elgar Publishing.
- Sirmans, G. S., MacDonald, L., and Macpherson, D. A. (2010). A meta-analysis of selling price and time-on-the-market. *Journal of Housing Research*, 19(2):139–152.
- Sirmans, G. S., MacDonald, L., Macpherson, D. A., and Zietz, E. N. (2006). The value of housing characteristics: a meta analysis. *The Journal of Real Estate Finance and Economics*, 33(3):215–240.
- Smith, T. and Wu, P. (2009). A spatio-temporal model of housing prices based on individual sales transactions over time. *Journal of Geographical Systems*, 11(4):333–355.
- Smith, T. E. (2009). Estimation bias in spatial models with strongly connected weight matrices. *Geographical Analysis*, 41(3):307–332.
- Stephan, F. F. (1934). Sampling errors and interpretations of social data ordered in time and space. *Journal of the American Statistical Association*, 29(185A):165–166.
- Sun, H., Tu, Y., and Yu, S.-M. (2005). A spatio-temporal autoregressive model for multi-unit residential market analysis*. *The Journal of Real Estate Finance and Economics*, 31(2):155–187.
- Tobler, W. (1979). *Cellular geography*, pages 379–386. Springer.
- Tu, Y., Yu, S. M., and Sun, H. (2004). Transaction-based office price indexes: A spatiotemporal modeling approach. *Real Estate Economics*, 32(2):297–328.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
- 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):1–63.

ANNEXE A

Articles considérant l'analyse en coupe transversale

Auteurs	Marché	Ville	Période
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
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
Cavailhès 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. (2013)	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
Kuethe (2012)	Résidentiel	Milwaukee (US)	2005 2005
Landry (2011)	Résidentiel	Savannah (US)	1990 1999