

A proposal of spatial lag operator: only the k-nearest neighbor (oknn).

Laurent Faucheux^a

March 23, 2019

Abstract

The text that follows is an extract of my PhD thesis, *A Multi-Scale Spatial General Equilibrium Model, Applied to the USA and France*. In this document, I describe GEMSE, a model whose aim is to investigate the interplays between aggregate and local dimensions of economic activity while quantifying GHG emissions associated to mobility within a spatially explicit framework. The model is based on the unification of Urban Economics and the New Economic Geography to model on multiple spatial scales the economic development of urban areas in interaction.

At some point of the model calibration, a subsection is devoted to the (spatial) econometric validation of the budget shares related to housing and transport services in workers expenditures. In parallel, it also contributes to the theoretical spatial-econometric debate on the appropriate choice of the spatial weight matrix and makes a new proposal for it. This proposal is implemented by using a numerical tool developed on purpose, namely PyOKNN. Applied to Greater Paris, the tool significantly identifies some interpretable key elements of the urban area's spatial autocorrelation structure, i.e the north-south and east-west dimensions of intra-mural Paris.

Key words: KW1, KW2, KW3, ..., KWN

JEL Code: LXX, LXX, LXX, LXX, LXX, ...

^aLO10 company, Email: laurent.faucheux@hotmail.fr

2.3 Econometric validation of the budget shares

A large amount of data has been gathered to calibrate GEMSE. Among those, note that housing rents are however not exogenously inputed at the calibration stage into GEMSE to initiate simulations, unlike, *e.g.* travel times. As explained in the housing related paragraph of subsection 2.2.3, housing rents are endogenerated by GEMSE, be it at the calibration or simulation stage. The assumption of calibration according to which the agglomeration is at its equilibrium state [see eq.(1.51)] allows for the deduction of housing rents in the entire urban area by only extrapolating those in the center.

In this section, observed (private vehicle) travel times in France in Paris are used in conjunction with observed housing rents to estimate the elasticities of utility with respect to accessibility, δ^a , and to housing services, δ^H . Put differently, we exploit the very nature of the transport/housing trade-off to estimate these elasticities. The aim is to assess whether the values of δ^H and δ^a that are chosen for France – respectively of 25% and 10% – are econometrically valid at the level of Paris. Note that this econometric validation could be declined into a spatial panel data applied to all urban areas of France and/or the USA. However, inclined not to resort to any kind of automatic selection techniques on the one hand and given that the approach that we follow requires careful handling on the other hand,²⁷ we decided to first concentrate on only one urban area, *i.e.* Paris.

Furthermore, this section takes part in the theoretical spatial-econometric (SE) debate on the right choice of the (so-denoted \mathbf{W}) spatial weight matrix (Anselin, 2007; Corrado and Fingleton, 2012; Elhorst et al., 2012; Gibbons and Overman, 2012) and makes a new proposal for it.

The proposal is implemented using a homemade Python library, named [PyOKNN](#). The [sourcecode](#) and [documentation](#) are available [online](#)^a and the package is easily installable via pip^b opening a session in your OS shell prompt and typing `pip install pyoknn`. This library is though of as a potential extension of another library called [PySAL](#) (Anselin and J. Rey, 2007).^c The lag operator is involved within the maximum likelihood estimator to conduct the estimation of the spatial ARIMA (SARIMA) model, which consists of the combination of the spatial error model (SEM), the spatial lag model (SAR) and the spatial moving average (SMA) model. It returns the same coefficients as those of PySAL when estimating monovariate [SAR](#) or [SEM](#) models, and allows for the computation of their multivariate version – where the terms *monovariate* and *multivariate*, refer to the number of parameters involved in spatial filters. Also, the confidence intervals that are computed in PySAL are analytically derived, while those of our library rely on numerical

²⁷Indeed, as will be explained subsequently, this study deals with the contingent determination of the structure of autoregression of housing rents.

hessian approximation and/or spatial bootstrap sampling.

^aAvailable at <https://github.com/lfaucheur/PyOKNN>

^bIt is a package management system used to install and manage software packages written in Python.

^cSee <http://pysal.readthedocs.io/en/latest/>.

Subsection 2.3.1 reviews the relevant literature, Subsection 2.3.2 offers the proposal and Subsection 2.3.4 applies this proposal to estimate δ^H and δ^a and assesses their representativity and coherence regarding the values that are chosen in Subsection 2.1.1.

2.3.1 Theoretical background

Fingleton (2009) and Corrado and Fingleton (2012) remind the analogies between temporal and spatial processes, at least when considering their lag operators. In the spatial econometric (SE) case, the lag operator is always explicitly involved *via* the use of a $n \times n$ matrix \mathbf{W} , where n is the number of interacting positions *in the scope of the econometric study only*.²⁸ The chosen space can be geographic, economic, social or of any other type. In the temporal case, which is seen as a space like no other given its inescapable anisotropic nature, the lag operator is in practice never explicitly considered. Any variable to lag, say, a $n \times 1$ vector \mathbf{y} , is formed over components that are beforehand sorted according to their position on the timeline.²⁹ This allows the lag-procedure to simply consist of offsetting down these components by a lag-determined number of rows, say, one row. In matrix terms, this offsetting procedure would be entirely equivalent to pre-multiplying an unsorted version of \mathbf{y} by a boolean $n \times n$ matrix \mathbf{H} with 1s indicating the immediate and unilateral proximity between temporal positions.

The so-structured data generating process (DGP) thus involves \mathbf{H} as primarily observed, *i.e.* with no restructuring hypothesis or transformation. For each lag, this provides the statistician with a straightforward parameter space definition, whose knowledge of the exact boundary is important, both for estimation and inference (Elhorst et al., 2012).

By opposition to the time series (TS) case, specifying \mathbf{W} involves a lot of arbitrariness. Apart from \mathbf{W} 's non-nilpotency,³⁰ these hypotheses deal with \mathbf{W} 's isotropy (Cressie, 1993) and finding \mathbf{W} 's true entrywise specification through a very large number of competing ones, be it functional³¹ or binary. Some famous entrywise specifications are the negative exponential function (Haggett, 1965), the inverse-distance function (Wilson, 1970), the combined distance-

²⁸Indeed, let's recall that in the major part of this thesis, n_j and n respectively stand for the number of firms in agglomeration j and in the country of interest. The GEMSE-consistent manner to denote the number of interacting positions would have been N_j . This is reminded in subsection 2.3.4 when starting to deal with the application.

²⁹The temporal lag operator illustrated by Corrado and Fingleton (2012) locates the most recent observation at the bottom of the to-be-lagged vector, *e.g.* at y_n in the case of a $n \times 1$ vector \mathbf{y} .

³⁰Which means that there is no permutation of the observational units that would make \mathbf{W} triangular, see Martellosio (2011).

³¹Whose parameters need to be estimated.

boundary function (Cliff and Ord, 1973) and the weighted logistic accessibility function (Bodson and Peeters, 1975).³² Binary weights specifications are either based on the k^{th} -nearest neighbor (knn), on the k^{th} -order of contiguity or on the radial distance. Then, to ensure the unique definition of any to-be-lagged variable in terms of the other variables of the model, \mathbf{W} is scaled depending on the choice one makes among three competing normalization techniques. The first one makes \mathbf{W} row-stochastic, but does not necessarily preserve its symmetry. The second one pre- and post-multiplies \mathbf{W} by the negative square root of a diagonal matrix reporting its row-totals (Cliff and Ord, 1973). The last one scales \mathbf{W} by its largest characteristic root (Elhorst, 2001).

But the choice of \mathbf{W} and of its transformation is not innocuous. For a maximum likelihood (ML) estimation to be consistent, the estimated spatial model must involve the true \mathbf{W} (Dogan, 2013; Lee, 2004). When dealing with autoregressive disturbances, both estimators ML and spatial generalized moments (GM) (Anselin, 2011; Arraiz et al., 2010; Drukker et al., 2013; Kelejian and Prucha, 2010)³³ theoretically base their knowledge of unobservable innovations upon the knowledge of \mathbf{W} . When facing endogeneity problems in non-autoregressive specifications and resorting to, *e.g.* Kelejian and Prucha (1999)'s generalized moments estimator (GM), the definition of the exogeneity constrains heavily relies on \mathbf{W} , which yields consistent and efficient estimations for sure, but potentially not with respect to the true DGP. If resorting to the instrumental variables (IV) method – in which space is conceived as providing ideal instruments (Das et al., 2003; Lee, 2003; Pinkse and Slade, 2010) –, the strength of instruments is far from being ensured with \mathbf{W} in its most common specification, *i.e.* whose lag consists of neighbors-averaging. Moreover, as discussed by Gibbons and Overman (2012), the inclusion of the product of higher powers of the spatial lag operator in the set of instruments is very likely to lead to a problem of collinearity, which in turn leads to the weaknesses of both identification and instruments. Finally, when computing LeSage and Pace (2009)'s total direct and indirect effects, the correctness of the true derivative of the regressand with respect to any spatially filtered³⁴ variable is a direct result of the correctness of \mathbf{W} .

In the following subsection, we propose a specification method for the spatial lag operator whose properties are as close as possible to that of its time series (TS) counterpart, *i.e.* usable as primarily observed without modifications. Nonetheless we follow Pinkse and Slade (2010, p.105)'s recommendation of developing *tools that are not simply extensions of familiar TS techniques to multiple dimensions*. We do so by proposing a specification-method that is fully grounded on the observation of the empirical characteristics of space, while minimizing as much as possible the set of hypotheses that are required. As clarified previously, this is from

³²Note that inverse-distance and negative exponential functions, as well as the continuum beyond and between the two, can be unified into the negative exponential of a modified Box-Cox transformation of any non-negative distance d . Formally, $e^{-\gamma(\min(\lambda,1)+d_\lambda)}$ with $d_\lambda = \frac{d^\lambda - 1}{\lambda}$ if $\lambda \neq 0$ and $d_\lambda = \ln d$ otherwise.

³³Among who, Arraiz et al. (2010) formalize moment conditions that allow for spatial lags in the dependent variable, the exogenous variables, and disturbances, the latter being assumed to be based on unknown-form heteroskedastic innovations.

³⁴The so-called spatial Cochrane-Orcutt style transformation.

the get-go in the way that \mathbf{W} is observed, which is made under hypotheses.

2.3.2 A proposal of lag operator

As Anselin (2010) states, Paelinck and Klaassen (1979) is the first to elaborate five rules designed to guide the formulation of spatial econometric models. With no elusive omission, two of these rules are (*ii*) the asymmetry in spatial relations and (*v*) the explicit modelling of space (topology) in spatial models.

These rules are actually straightforwardly violated by some entrywise specifications of \mathbf{W} . The negative exponential function, the distances-inverse function or binary weights based on radial distance are prominent examples. Indeed, when formalizing the spatial relationship between two given spatial units i and j , these purely distance-based entrywise specifications fall short of formalizing asymmetric relations for the simple reason that $d_{ij} = d_{ji}$ is traditionally assumed. Incidentally, preserving this symmetry when making \mathbf{W} row-stochastic is seen as desirable.³⁵

These specifications furthermore constrain the explicit modelling of space upon an arbitrarily parameterized functional form. From the positions' standpoint, this inevitably instills non-neutral and rigid *kernels* about the masses of autoregressive effects that flow from them over their neighbors. Figure 2.5 illustrates this point upon Anselin (1988)'s Columbus, Ohio,³⁶ polygon 25,³⁷ respectively with the distance inverse function, $d_i^{-\gamma}$, the exponential negative function, $e^{-\gamma d_i}$, and the first-order contiguity, c_i^{1st} . From the point of view of neighbors – over

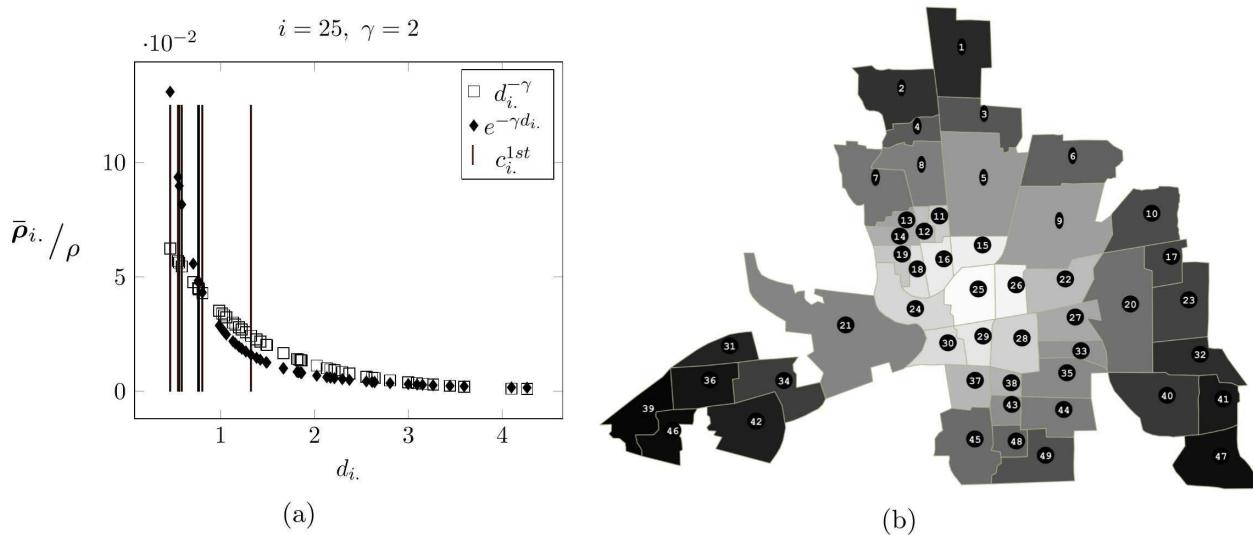


Figure 2.5: Columbus polygon 25's (a) normalized weights and (b) mapped distances from it

all positions – these kernels are very likely to be heterogeneous since for each order of proximity,

³⁵This is seen as desirable since having \mathbf{W} symmetric also means not having its eigenvalues in the complex domain, \mathbb{C} , which keeps the pro-stationarity parameter space definition entirely included in \mathbb{R} .

³⁶This data set is used since it is one of the most famous demonstration data sets in Spatial Econometric.

³⁷The polygon is selected for its central position in the lattice.

it is uncommon with $\{(i, j) : w_{ij} > 0\}$ to get $\bar{\rho}_{ij} = \rho$ as in the TS case, since $\frac{\bar{\rho}_{ij}}{\rho} = \frac{\mathbf{w}_{i\cdot}}{\sum_{j=1}^n w_{ij}}$, where ρ is a scalar autoregressive parameter and $\mathbf{w}_{i\cdot}$ is the i -th row of the weights matrix \mathbf{W} . Note that even flat/uniform kernels may instill non-neutral heterogeneity as long as the numbers of neighbors differ from one position to another. If the number of neighbors is identical across positions, still remains the impossibility to distinguish one order of proximity from the other. Finally, all entrywise specifications that imply non-binary weights and/or more than one nonzero weight per row, unlike *e.g.* a 1-nearest neighbor matrix,³⁸ are concerned by this problem.

It is worthwhile wondering why autoregressive effects should flow in that so-structured and smooth manner from polygon 25 over its neighborhood?³⁹ What about decomposing this outflow in an atomic fashion, *e.g.* by considering each order of proximity in separated correlation structures rather than in one piece? Avoiding this non-neutral kernel is the main *raison d'être* of the proposal that follows, baptized *k-nearest neighbor only (oknn)*.

First, denote by $\mathbb{D}(n)$ the set of all $n \times n$ distance matrices and by $\mathbf{D} \in \mathbb{D}(n)$ a matrix whose typical element, \mathbf{d}_{ij} , is the distance from any spatial unit $i = 1, \dots, n$ to any spatial unit j of the lattice. Second, for each order of autoregression, say, for each $k = 1, \dots, n - 1$, denote by \mathbf{D}_k an $n \times n$ matrix whose typical element $\mathbf{d}_{ij,k}$ equates \mathbf{d}_{ij} instead of zero if \mathbf{d}_{ij} is the k th smallest distance from the spatial unit i , *i.e.* $\mathbf{d}_{ij,k} = \mathbf{d}_{ij}$ if $r_i(\mathbf{d}_{ij}) = k$ and $\mathbf{d}_{ij,k} = 0$ otherwise, where r_i is a function that ranks distances from the spatial unit i to all units $j \neq i$. Furthermore, write $\mathbf{d}_{i\cdot,k}$ for the i th row of \mathbf{D}_k , $\mathcal{I}_{\mathbf{D}_k}$ for the set of all neighbors that are considered through \mathbf{D}_k at rank k and $\mathcal{I}_{i,\mathbf{D}_k}$ or $\mathcal{I}_{\mathbf{d}_{i\cdot,k}}$ for the set of neighbors of the spatial unit i that are considered through the i th line of \mathbf{D}_k at rank k , where $\mathcal{I}_{\mathbf{D}_k} = \bigcup_{i=1}^n \mathcal{I}_{\mathbf{d}_{i\cdot,k}}$. Note that any $n \times n$ distance matrix $\mathbf{D} \in \mathbb{D}(n)$ can be seen as coordinates that characterize the corresponding lattice by locating it in $\mathbb{R}_{\geq 0}^{n(n-1)/2}$.^a

Assumption 1. There exists a surjective triple $(\mathbb{R}_{\geq 0}^{n(n-1)/2}, D, \mathbb{R}_{\geq 0})$, *i.e.* a surjective function, D , that takes $\mathbf{D} \in \mathbb{D}(n)$ as argument and increases indefinitely in the lattice irregularity.

Assumption 2. For $k = 1, \dots, n - 1$, any spatial unit $i = 1, \dots, n$ always has at least 1 k th neighbor no matter how far they find themselves from each other, *i.e.* $\text{card}(\mathcal{I}_{i,\mathbf{D}_k}) \geq 1 \forall \mathbf{D} \in \mathbb{D}(n)$. This means that any spatial unit i has exactly one k th neighbor when the lattice is perfectly irregular, *i.e.* $\text{card}(\mathcal{I}_{i,\mathbf{D}_k}) \rightarrow 1$ as $D(\mathbf{D}) \rightarrow \infty$.

Proposition 1. For $k = 1, \dots, n - 1$, $\text{card}(\mathcal{I}_{\mathbf{D}_k}) = O(n)$ as $D(\mathbf{D}) \rightarrow \infty$.

³⁸Another name of entrywise specification that represents the same object is the binary contiguity matrix of the first-order restricted to 1 neighbor.

³⁹Tobler's first law of geography is a the very first answer to this question.

Proof of Proposition 1. For $k = 1, \dots, n - 1$, by definition $\mathcal{I}_{\mathbf{D}_k} = \bigcup_{i=1}^n \mathcal{I}_{i,\mathbf{D}_k}$ and with assumption 2, it follows that $\text{card}(\mathcal{I}_{\mathbf{D}_k}) = O(n)$ as $D(\mathbf{D}) \rightarrow \infty$. \square

Assumption 3. \mathbf{D} is such that $\text{card}(\mathcal{I}_{\mathbf{D}_k}) = O(n)$ for $k = 1, \dots, n - 1$.

This assumption is mandatory to characterize our proposal of lag operator, whose rational is the irregularity of the lattice under consideration.

Proposition 2. For $k = 1, \dots, n - 1$, \mathbf{D}_k contains exactly n non-zero elements.

Proof of Proposition 2. From assumptions 2 and 3, it follows that any position $i \in \llbracket 1; n \rrbracket$ has exactly one k th neighbor for $k = 1, \dots, n - 1$. Put differently, it follows that $\forall i \in \llbracket 1; n \rrbracket$, $\text{card}(\mathcal{I}_{\mathbf{d}_{i,k}}) = 1$ for $k \in \llbracket 1; n \rrbracket$. Thus $\sum_{i=1}^n \text{card}(\mathcal{I}_{\mathbf{d}_{i,k}}) = n$. \square

To ultimately define \mathbf{W}_k from \mathbf{D}_k for each $k = 1, \dots, n - 1$, simply replace each \mathbf{D}_k 's nonzero typical element, $\mathbf{d}_{ij,k}$, by 1. It follows that the assumption about the diagonal elements of the \mathbf{W}_k s is redundant since directly attributable to the distance-matrices intrinsic zero diagonal elements.

Proposition 3. For $k = 1, \dots, n - 1$, \mathbf{W}_k is row-stochastic.

Proof of Proposition 3. From assumption 2, it follows that any position $i \in \llbracket 1; n \rrbracket$ has exactly one k th neighbor for $k = 1, \dots, n - 1$, *i.e.* $\text{card}(\mathcal{I}_{\mathbf{d}_{i,k}}) = 1$. Thus for $k = 1, \dots, n - 1$, \mathbf{W}_k has exactly one 1 per row, it is row-stochastic. \square

Proposition 4. For $k = 1, \dots, n - 1$, no \mathbf{W}_k can be a linear combination of the others.

Proof of Proposition 4. From assumptions 2 and 3, it follows that any position $i \in \llbracket 1; n \rrbracket$ has exactly $n - 1$ different neighbors. Put differently, from these two assumptions it follows that $\text{card} \bigcup_{k=1}^{n-1} \mathcal{I}_{\mathbf{w}_{i,k}} = \text{card}(\{\mathbf{w}_{i,k} : k \in \llbracket 1; n \rrbracket\}) = n - 1$, *i.e.* a k th neighbor cannot also be a $(k + j)$ th neighbor, where $j \in \llbracket -k + 1; n - k \rrbracket \setminus \{0\}$. Since the (row-) components of \mathbf{W}_k cannot be a linear combination of themselves at different neighborhood orders, so is \mathbf{W}_k . \square

^aWhere $\frac{n(n-1)}{2}$ is the exact number of degrees of freedom over which a lattice can be uniquely defined, given that \mathbf{D} , as a distance matrix, is intrinsically Hollow and symmetric.

As far as we know, whereas the oknn specification of \mathbf{W}_k is the strict spatial counterpart of the k -order TS lag operator, \mathbf{H}^k , it had surprisingly never been proposed. The likely reason for this fact is the usual assumption of regular lattice, on which the autoregression structure superimposes.⁴⁰ Frequently seen as an issue, the irregularity of the lattice is the rational for this specification. Moreover, in realistic spatial configurations, the lattice regularity is the ex-

⁴⁰In the regular lattice case, spatial lag operators differ from TS's by locating more than one neighbor for a given separating distance.

ception rather than the rule.⁴¹

2.3.3 Involving oknn in spatial ARIMA process

The oknn specification for \mathbf{W}_k possesses a high level of similarities with its TS counterpart, previously denoted as \mathbf{H}^k , *e.g.* they are non-normalized since originally row-stochastic, they are boolean permutation matrix and do not overlap with other lag orders. However, a major difference persists: the absence of the (linear) algebraic relationship that links lag orders between them, *e.g.* to the first one within an integer power transformation such that $\mathbf{H}^k = (\mathbf{H}_1)^k$ for $k = 1, \dots, n - 1$.⁴² A direct consequence of this is the non-trivial choice of studying high-order polynomial in k spatial weights matrices \mathbf{W}_k for $k = 1, \dots, n - 1$. To illustrate this point, consider a process with two orders of partial integration in \mathbf{y} . In the TS case, the second-order polynomial has the property that $(\mathbf{I} - \rho_1 \mathbf{H}^1)(\mathbf{I} - \rho_2 \mathbf{H}^2)\mathbf{y} = (\mathbf{I} - \rho_1 \mathbf{H}^1 - \rho_2 \mathbf{H}^2 + \rho_1 \rho_2 \mathbf{H}^3)\mathbf{y}$.⁴³ A contrario, in the SE case, it is (astronomically) unlikely that $\mathbf{W}_1 \mathbf{W}_2 = \mathbf{W}_3$. As stated by Elhorst et al. (2012), "a logical implication of this view of modeling spatial dependence [...] implies that extending the first order model to include more than one spatial weights matrix requires that we consider the cross product term". However, in that case, what about considering the (explosive) combinatorial of the cross product with other matrices as well, *i.e.* the transposed cross product, the cross product of $\mathbf{W}_1 \mathbf{W}_2$ with $\mathbf{W}_1, \mathbf{W}_2$, and so on?⁴⁴ The questions that this poses are not dealt with in this study and remain a topic for future research.

One advantage of the oknn specification is that there is no need when involving *only one lag* to consider avoiding singularity or explosive processes by altering the parameter space definition. For example, if one considers a first order spatial autoregressive process (SAR), the pro-stationarity parameter space of the underlying autoregression coefficient –expressed in terms of the minimum and maximum eigenvalues of \mathbf{W}_k , respectively ω_{\min} and ω_{\max} – is $[\omega_{\min}^{-1}, \omega_{\max}]$. Note that for row-stochastic matrices one has $\omega_{\max}^{-1} = 1$, but no general result holds for ω_{\min}^{-1} (Anselin, 1982). However, when in addition to being row-stochastic, the lag operator is a permutation matrix as in the TS or oknn case, its eigenvalues necessarily lie on the unit circle,⁴⁴ which means that one actually has a general result in the oknn case, that is that $\omega_{\min}^{-1} \leq -1$. In the case of *multiple lags*, say, m , the definition of the admissible parameter space within \mathbb{R}^m is less straightforward. As pointed out by Elhorst et al. (2012), the naive adoption of parameter space restricted such that $\sum_{i=1}^m |\rho_i| < 1$ "proves to be too restrictive".

⁴¹Which is the only reason why \mathbf{W}_k is k -subscripted while \mathbf{H}^k is k -superscripted since the beginning of this work.

⁴²Putting aside the discussion about testing that $\rho_1 \rho_2 + \rho_3 = 0$.

⁴³Note that in the TS case, the number of possible unique combinations is finite, *i.e.* it is n , with $(\mathbf{H}_1)^n = \mathbf{0}_{n,n}$

⁴⁴A permutation matrix can always be expressed as a product of independent rotation matrices with eigenvalues that lie on the unit circle, $e^{2ik\pi/n} \forall k \in \mathbb{Z}$, and thus whose non-imaginary part necessarily lies on the (rational) segment $[-1, 1]$. Note that we consider only the real part of eigenvalues since only those influence spatial filters' singularity (LeSage and Pace, 2009).

It is nevertheless far more desirable than the opposite situation in which this naive adoption would be not restrictive enough. Thus, this is both for the sake of caution and simplicity that we follow Lee and Liu (2010) and Badinger and Egger (2011) by making the assumption that the sum of the absolute values of the parameters (involved in the same spatial filter) should be less than one. Making this assumption also has the positive consequence of ensuring that each parameter, taken individually, complies with its own one-dimensional space definition.

With a little loss of generality since we neglect k -order polynomials, we put the focus on k -order spatial processes and we directly transpose in space the three-stage modeling approach of Box and Jenkins (1976) that consists of (*i*) identifying and selecting the model,⁴⁵ (*ii*) estimating the parameters and (*iii*) checking the model.

Know however that we do not use the usual TS notation with both nonseasonal and seasonal factors, *i.e.* ARIMA(p, d, q) \times (P, Q, D)_s,⁴⁶. By analogy, we rather denote any model by using sets of lags, as ARIMA({1, ..., p, P}, {1, ..., d, D}, {1, ..., q, Q}). The reason for this notation choice is intended to be informative. As explained by Anselin (2002a, p.253-254), spatial pro-stationarity differences can only be performed partially compared to the TS case. Indeed, these differences must include autoregressive parameters – thus not implicitly set to 1 as in the TS case – so as to avoid singularities when handling models in their reduced form. Anselin mentions in addition that, unlike the TS case, any spatial filter estimation must be carried out jointly with that of the other model parameters. Put in technical terms, this means that during estimations, the two TS steps of stationarity differencing and model identification are merged together.

The development that follows is clarified in Section 2.B. We initiate the discussion with the traditional linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \quad (2.23)$$

with \mathbf{y} as a $n \times 1$ vector of observations on the endogenous variable, \mathbf{X} as an $n \times k$ matrix of observations on exogenous variables, $\boldsymbol{\beta}$ as a $k \times 1$ vector of coefficients and \mathbf{v} as a homoskedastic and non-autocorrelated $n \times 1$ vector of disturbances, such that $E[\mathbf{v}\mathbf{v}'] = \sigma_v^2 \mathbf{I}$.

If the disturbance terms of the model are more forcefully assumed i.i.d. normal with mean zero, *i.e.* $\mathbf{v}|\mathbf{X} \sim \mathcal{N}(0, \sigma_v^2 \mathbf{I})$, its multivariate density is

$$\mathcal{L}(\mathbf{v}) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\mathbf{v}'\mathbf{v})} \quad (2.24)$$

The likelihood for \mathbf{y} conditional on \mathbf{X} is then

$$\mathcal{L}(\mathbf{y}|\mathbf{X}) = \mathcal{L}(\mathbf{v}) \left| \frac{\partial \mathbf{v}}{\partial \mathbf{y}} \right| \quad (2.25)$$

⁴⁵It is interesting to note that this phase may as well easily be thought of as the identification/selection procedure of the full autocorrelation structure of the variables involved in the data-generating process.

⁴⁶The terms (p, d, q) and $(P, D, Q)_s$ respectively give the orders of the nonseasonal and seasonal parts, where s is the number of observations in a seasonal cycle, *e.g.* 12 for monthly series, 4 for quarterly series, 7 for daily series with day-of-week effects.

where $|(\partial\mathbf{v}/\partial\mathbf{y})|$ is the absolute value of the determinant formed from the $n \times n$ matrix of partial derivatives of the elements of \mathbf{v} with respect to the elements of \mathbf{y} .

One may then think that eq.(2.25) should be augmented to entail different types of spatial effects. Considering these effects can be done through different spatial model specifications, the most notorious (and elementary) of which are the spatial autoregressive model (SAR), the spatial error model (SEM), the spatial moving average model (SMA) and *different combinations of those*.⁴⁷ However the way of combining them is not unique. This is why, as mentioned previously, the chosen combination and its analytic development is driven by the first phase of identification/selection. We recall that even if the step of pro-stationarity differencing presently initiates the selection phase – as classically performed in the TS case –, it will always be re-carried out jointly with the estimation of the other model parameters. This is clarified in the following respects.

Identifying the order and the structure of differencing

In the TS case, the first step is about making sure that the variables are stationary. In the SE context, this means assuming that the data-generating process (DGP) can be specified as a SEM, or less particularly, as a spatial Durbin model (SDM). Indeed, the SEM is a special case of SDM in which both \mathbf{y} and \mathbf{X} are assumed to be (partially) identically integrated, *i.e.* the difference structure is assumed to be the same on both the exogenous and endogenous variables (Anselin, 1980; Elhorst and Vega, 2013; Gibbons and Overman, 2012). However, the declination of the SEM into a SDM within the search of the order of differencing is not undertaken in this study.⁴⁸ Moreover, note that this subsection is – for the sake of simplicity – more about *differencing structure* than about differencing order. Actually, we chose to restrict the latter to the binary case of differencing or not differencing. Put differently, we chose to restrict the latter to contingently consider *how relatively high* or *how absolutely high* model's variables are over the urban space. Thus, still remains to determine the spatial structure of this first difference, if any. For the sake of clarity, let first denote

$$\begin{aligned}\mathbf{G}(\boldsymbol{\gamma}) &= \sum_{i \in \mathcal{I}_{\boldsymbol{\gamma}}} \gamma_i \mathbf{W}_i \\ \boldsymbol{\Gamma}(\boldsymbol{\gamma}) &= \mathbf{I} - \mathbf{G}(\boldsymbol{\gamma}) \\ \mathbf{y}_{\boldsymbol{\gamma}} &= \boldsymbol{\Gamma}(\boldsymbol{\gamma}) \mathbf{y} \\ \mathbf{X}_{\boldsymbol{\gamma}} &= \boldsymbol{\Gamma}(\boldsymbol{\gamma}) \mathbf{X} \\ \boldsymbol{\Omega}_{\boldsymbol{\gamma}} &= \sigma_u^2 \left(\boldsymbol{\Gamma}' \boldsymbol{\Gamma} \right)^{-1}\end{aligned}$$

⁴⁷At the combo-top of which one has the general nesting spatial model (GNS). See Elhorst and Vega (2013, p.24, Fig.1) for a general-to-specific three-like comparison of different spatial econometric model specifications. Incidentally, the spatial moving average model is a notable absentee.

⁴⁸The declination could be performed one step further by considering how a SDM boils down to a spatial lag (only) of \mathbf{X} (SLX) model. Although this is not related to the notion of differencing.

where \mathcal{I}_γ is the set of lag orders that are considered for estimation, $\gamma_i \forall i \in \mathcal{I}_\gamma$ is the i -th spatial autoregressive parameter involved in the auto-generation of \mathbf{v} , $\boldsymbol{\gamma}$ is the $\text{Card}(\mathcal{I}_\gamma) \times 1$ vector formed over these autoregressive parameters and \mathbf{W}_i is the i -associated $n \times n$ spatial oknn-lag operator.

Henceforth it is assumed that $\mathbf{v} = \mathbf{G}(\boldsymbol{\gamma})\mathbf{v} + \mathbf{u}$ with $\mathbf{u} \sim \mathcal{N}(0, \sigma_u^2 \mathbf{I})$ and $\mathbf{v} \sim \mathcal{N}(0, \boldsymbol{\Omega}_\gamma)$ [see Subsection 2.B.1 for details]. This leads to rearrange eq.(2.23) as

$$\mathbf{y}_\gamma = \mathbf{X}_\gamma \boldsymbol{\beta} + \mathbf{u} \quad (2.26)$$

Substituting closed form solutions from the first order conditions for the parameters $\boldsymbol{\beta}$ and σ_u^2 leads eq.(2.25) to be concentrated on $\boldsymbol{\gamma}$ as follows

$$\mathcal{L}(\boldsymbol{\gamma}) = \left(\frac{2\pi e}{n} \mathbf{u}' \mathbf{u} \right)^{-\frac{n}{2}} |\boldsymbol{\Gamma}(\boldsymbol{\gamma})| \quad (2.27)$$

Recalling that eq.(2.27) reduces to eq.(2.25) for $\mathcal{I}_\gamma = \emptyset$.

Identifying the numbers of AR or MA terms

One may then have reasons to think that, specified as in eq.(2.26) – or as in eq.(2.23) if $\mathcal{I}_\gamma = \emptyset$ –, one omits observable and/or non-observable characteristics at work in the DGP. On the one hand, one of those observable characteristics is very likely to be the long-distance auto-determining nature of the spatially-filtered regressand, \mathbf{y}_γ . On the other hand, if the long-distance auto-regressive nature of the error vector is already considered by the SEM specification, a non-observable characteristic that is likely to remain is its short-distance autocorrelation. For the sake of clarity relatively to the SAR terms, let denote

$$\begin{aligned} \mathbf{Q}(\boldsymbol{\rho}) &= \sum_{i \in \mathcal{I}_\rho} \rho_i \mathbf{W}_i \\ \mathbf{P}(\boldsymbol{\rho}) &= \mathbf{I} - \mathbf{Q}(\boldsymbol{\rho}) \\ \mathbf{y}_{\gamma, \boldsymbol{\rho}} &= \mathbf{P}(\boldsymbol{\rho}) \mathbf{y}_\gamma \end{aligned}$$

where \mathcal{I}_ρ is the set of lag orders that are considered for estimation, $\rho_i \forall i \in \mathcal{I}_\rho$ is the i -th spatial autoregressive parameter involved in the auto-generation of \mathbf{y}_γ and $\boldsymbol{\rho}$ is the $\text{Card}(\mathcal{I}_\rho) \times 1$ vector formed over these autoregressive parameters.

Identically for the SMA terms, let denote

$$\begin{aligned}
\mathbf{L}(\boldsymbol{\lambda}) &= \sum_{i \in \mathcal{I}_{\boldsymbol{\lambda}}} \lambda_i \mathbf{W}_i \\
\boldsymbol{\Lambda}(\boldsymbol{\lambda}) &= \mathbf{I} + \mathbf{L}(\boldsymbol{\lambda}) \\
\mathbf{y}_{\gamma, \rho, \lambda} &= \boldsymbol{\Lambda}(\boldsymbol{\lambda})^{-1} \mathbf{y}_{\gamma, \rho} \\
\mathbf{X}_{\gamma, \rho} &= \boldsymbol{\Lambda}(\boldsymbol{\lambda})^{-1} \mathbf{X}_{\gamma} \\
\boldsymbol{\Omega}_{\lambda} &= \sigma_r^2 \boldsymbol{\Lambda} \boldsymbol{\Lambda}' \\
\boldsymbol{\Omega}_{\gamma, \rho, \lambda} &= \mathbf{Q} \text{cov}[\mathbf{y}_{\gamma}, \mathbf{y}_{\gamma}] \mathbf{Q}' \\
&\quad + 2 \mathbf{Q} \mathbb{E}[\mathbf{y}_{\gamma} \mathbf{r}'] \boldsymbol{\Lambda}' \\
&\quad + \boldsymbol{\Omega}_{\lambda}
\end{aligned}$$

where $\mathcal{I}_{\boldsymbol{\lambda}}$ is the set of lag orders that are considered for estimation, $\lambda_i \forall i \in \mathcal{I}_{\boldsymbol{\lambda}}$ is the i -th spatial parameter involved in the local auto-regression of residuals and $\boldsymbol{\lambda}$ is the $\text{Card}(\mathcal{I}_{\boldsymbol{\lambda}}) \times 1$ vector formed over these autoregressive parameters. Note that the MA filter, $\boldsymbol{\Lambda}(\boldsymbol{\lambda})$, is expressed following Anselin as a positive sum.⁴⁹

Without loss of generalities, considering simultaneously the SAR and SMA specific terms while still opting for a $\text{SEM}(\mathcal{I}_{\gamma})$ is equivalent to suspecting that \mathbf{u} entails $\mathbf{Q}(\boldsymbol{\rho}) \mathbf{y}_{\gamma} + \boldsymbol{\Lambda}(\boldsymbol{\lambda}) \mathbf{r}$ with $\mathbf{r} \sim \mathcal{N}(0, \sigma_r^2 \mathbf{I})$, $\mathbf{u} \sim \mathcal{N}(0, \boldsymbol{\Omega}_{\gamma, \rho, \lambda})$ and $\mathbf{v} \sim \mathcal{N}(0, \boldsymbol{\Gamma}^{-1} \boldsymbol{\Omega}_{\gamma, \rho, \lambda} \boldsymbol{\Gamma}^{-1}')$ [see Subsection 2.B.2 for details]. Note that the inclusion of SAR terms is also related to the identification of the structure of partial differencing, since it relates to only considering \mathbf{y}_{γ} independently from the other variables of the model. As pointed out by Anselin (2002b, p.254), "this can be interpreted as a way to clean \mathbf{y}_{γ} of the effects of spatial correlation, while maintaining the [...] estimates for $\boldsymbol{\beta}$ ". This leads to rearrange eq.(2.26) as

$$\mathbf{y}_{\gamma, \rho, \lambda} = \mathbf{X}_{\gamma, \lambda} \boldsymbol{\beta} + \mathbf{r} \tag{2.28}$$

and eq.(2.27) to be concentrated on $\boldsymbol{\gamma}$, $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$ as follows

$$\mathcal{L}(\boldsymbol{\theta}) = \left(\frac{2\pi e}{n} \mathbf{r}' \mathbf{r} \right)^{-\frac{n}{2}} \frac{|\boldsymbol{\Gamma}(\boldsymbol{\gamma})| |\mathbf{P}(\boldsymbol{\rho})|}{|\boldsymbol{\Lambda}(\boldsymbol{\lambda})|} \tag{2.29}$$

where $\boldsymbol{\theta}$ denotes the $(\text{Card}(\mathcal{I}_{\gamma}) + \text{Card}(\mathcal{I}_{\rho}) + \text{Card}(\mathcal{I}_{\lambda})) \times 1$ hyperparameter that vertically stacks $\boldsymbol{\gamma}$, $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$, such that $\boldsymbol{\theta} = [\boldsymbol{\gamma}', \boldsymbol{\rho}', \boldsymbol{\lambda}']'$.

⁴⁹Indeed, probably following the convention introduced by Box and Jenkins, it is very common to read authors who express it exhibiting a minus sign. We prefer the form with a positive sum since it reminds a kind of truncated Leontief inverse, from which the interpretation in terms of auto-regression order is (subjectively) direct.

Dispersion for the parameters

In addition to the statistical inference required during the phase of selection/identification of the model, we recall the need to conduct inference on the budget shares, $\widehat{\delta}^H$ and $\widehat{\delta}^a$, that are set in GEMSE. The estimated values of these parameters derive from the estimated components of $\widehat{\beta}$ – whose functional relations are explicated in subsection 2.3.4.

The models that are subject to selection in the present work are likely to involve a large number of parameters whose distributions, probably not symmetrical, are cumbersome to derive analytically. This is why in addition to the (normal-approximation-based) observed confidence intervals,⁵⁰ (non-adjusted and adjusted) bootstrap percentile intervals are provided. However, the existence of fixed spatial weight matrices prohibits the use of traditional bootstrapping methods.⁵¹ So as to compute (normal approximation or percentile-based) confidence intervals for all the parameters, be them derived like the budget shares, we use a special case of bootstrap method, namely Lin et al. (2007)'s hybrid version of residual-based recursive wild bootstrap.⁵² This method is particularly appropriate since it (*i*) "accounts for fixed spatial structure and heteroscedasticity of unknown form in the data" and (*ii*) "can be used for model identification (pre-test) and diagnostic checking (post-test) of a spatial econometric model". As mentioned above, non-adjusted percentile intervals as well as bias-corrected and accelerated (BCa) percentile intervals (Efron and Tibshirani, 1993) are provided. An issue-based summary of the calculation methods of all the intervals that are presented in this work follows.

Confidence intervals that are based on the observed Fisher information matrix rely on the symmetry of the distribution of the parameters under question. The (observed) standard errors and confidence bounds of these parameters, respectively denoted by $\text{se}[\widehat{\Theta}]^{\text{obs}}$ and $\mathbf{c}[\alpha]_{\pm}^{\text{obs}, \Theta}$,^a are computed as

$$\text{se}[\widehat{\Theta}]^{\text{obs}} = \text{diag} \left[\left(-\frac{\partial \ln \mathcal{L}^{\text{full}}}{\partial^2 \Theta} \Big|_{\widehat{\Theta}} \right)^{-1} \right]^{1/2}; \quad \mathbf{c}[\alpha]_{\pm}^{\text{obs}, \Theta} = \widehat{\Theta} \pm \Psi_{n-k}^{-1}[\alpha/2] \text{se}[\widehat{\Theta}]^{\text{obs}}$$

where $\widehat{\Theta} = [\widehat{\beta}', \widehat{\theta}', \widehat{\sigma}_r^2]'$, α is the chosen probability of making type I error and Ψ_{n-k} is the student distribution function with $n - k$ degrees of freedom. To compute the observed standard errors of $\widehat{\delta}^H$ and $\widehat{\delta}^a$, one needs to redefine $\ln \mathcal{L}^{\text{full}}$ in terms of these two parameters

⁵⁰The term "observed" stands for – as it reads in open-source codes –, the computation of the log-likelihood Fisher Information matrix evaluated/observed at its estimated maximum.

⁵¹Seminarily without dependence (Bradley, 1979) or with (nilpotent anisotropic) temporal dependence by moving blocks (Kunsch, 1989). See Gonçalves and Politis (2011) for a short but very instructive review of bootstrap methods for TS.

⁵²The way to mimick heteroskedasticity of unknown form is detailed and tested in Davidson and Flachaire (2008).

such that (i) $\boldsymbol{\delta} = f(\boldsymbol{\beta})$, where f is a vector isomorphic function, (ii) $\boldsymbol{\Xi} = [f^{-1}(\boldsymbol{\delta})', \boldsymbol{\theta}', \sigma_r^2]'$, (iii) $\boldsymbol{\xi} = [\boldsymbol{\delta}', \boldsymbol{\theta}', \sigma_r^2]'$ and (iv) $\widehat{\boldsymbol{\xi}} = [f(\widehat{\boldsymbol{\beta}})', \widehat{\boldsymbol{\theta}}', \widehat{\sigma}_r^2]'$. These observed standard errors and confidence bounds, respectively denoted by $\text{se}[\widehat{\boldsymbol{\Xi}}]^{\text{obs}}$ and $\mathbf{c}[\alpha]_{\pm}^{\text{obs}, \boldsymbol{\Xi}}$, are

$$\text{se}[\widehat{\boldsymbol{\Xi}}]^{\text{obs}} = \text{diag} \left[\left(-\frac{\partial \ln \mathcal{L}^{\text{full}}(\boldsymbol{\Xi})}{\partial^2 \boldsymbol{\xi}} \Big|_{\widehat{\boldsymbol{\xi}}} \right)^{-1} \right]^{1/2} ; \quad \mathbf{c}[\alpha]_{\pm}^{\text{obs}, \boldsymbol{\Xi}} = \widehat{\boldsymbol{\Xi}} \pm \Psi_{n-k}^{-1}[\alpha/2] \text{se}[\widehat{\boldsymbol{\Xi}}]^{\text{obs}}$$

Unlike the above two approaches, mutually exclusive, confidence intervals that are based on bootstrap sampling distributions allow for the parallel estimation of the confidence bounds of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$. Define $\widehat{\boldsymbol{\Theta}}^{\text{all}}$ as a vector that vertically stacks all the parameters, be them directly estimated like $\boldsymbol{\beta}$ or indirectly like $\boldsymbol{\delta}$, such that $\widehat{\boldsymbol{\Theta}}^{\text{all}} = [\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\delta}}', \widehat{\boldsymbol{\theta}}', \widehat{\sigma}_r^2]'$. BCa percentile intervals are used to turn the bootstrap sampling distribution of $\widehat{\boldsymbol{\Theta}}^{\text{all}}$ into its analogous unbiased and constant variance version. The transformation implies two independent correction factors, commonly denoted by \mathbf{z} and \mathbf{a} ,^b which consider respectively the bias and the skewness of the bootstrap sampling distribution. \mathbf{z} and \mathbf{a} are defined as

$$\mathbf{z} = \Phi^{-1} \left(\frac{1}{r} \# \left(\widehat{\boldsymbol{\Theta}}_b^{\text{all}} < \widehat{\boldsymbol{\Theta}}^{\text{all}} \right) \right) ; \quad \mathbf{a} = \frac{1}{6} \frac{\sum_{i=1}^n (\widehat{\boldsymbol{\Theta}}_{-i}^{\text{all}} - \overline{\boldsymbol{\Theta}}^{\text{all}})^3}{\left(\sum_{i=1}^n (\widehat{\boldsymbol{\Theta}}_{-i}^{\text{all}} - \overline{\boldsymbol{\Theta}}^{\text{all}})^2 \right)^{3/2}}$$

where, on the one hand, b , r and Φ respectively stand for the bootstrap replication index, the total number of replications and the cumulative standard normal distribution function and, on the other hand, $\widehat{\boldsymbol{\Theta}}_{-i}^{\text{all}} - \overline{\boldsymbol{\Theta}}^{\text{all}}$ is the mean deviation of the i th jackknife estimate of (the true) $\boldsymbol{\Theta}^{\text{all}}$. Note that, if any,^c oknn matrices involved during the n jackknife estimation procedures are resampled as well by (i) removing their i th row and column and (ii) computing the corresponding oknn auto-correlation structure. Finally, define the correcting percentile (confidence interval) function, $\mathbf{c}[\alpha]_{\pm}^{\text{BCa}}$, as

$$\mathbf{c}[\alpha]_{\pm}^{\text{BCa}} = \mathcal{Q} \left(\Phi \left(\mathbf{z} + \frac{\mathbf{z} \pm \Phi^{-1}(\alpha/2)}{1 - \mathbf{a}(\mathbf{z} \pm \Phi^{-1}(\alpha/2))} \right) \right)$$

where \mathcal{Q} is the empirical quantile function that returns $\mathbf{c}[\alpha]_-$ or $\mathbf{c}[\alpha]_+$ such that, respectively, $\Pr(\boldsymbol{\Theta}^{\text{all}} < \mathbf{c}[\alpha]_-) = \alpha/2$ or $\Pr(\mathbf{c}[\alpha]_+ < \boldsymbol{\Theta}^{\text{all}}) = \alpha/2$, depending on whether or not $\mathbf{c}[\alpha]_{\pm}$ are BCa-based.

Thus, the confidence and percentile intervals that are computed in this work are four in number, above denoted as $\mathbf{c}[\alpha]_{\pm}^{\text{obs}, \boldsymbol{\Theta}}$, $\mathbf{c}[\alpha]_{\pm}^{\text{obs}, \boldsymbol{\Xi}}$, $\mathbf{c}[\alpha]_{\pm}$ and $\mathbf{c}[\alpha]_{\pm}^{\text{BCa}}$. Those will be presented for each coefficient in table format and referred respectively to as $D1$, $D2$, $D3$ and $D4$,

with D standing for (underlying) *Distribution*. Moreover, note that the definitions of the (BCa and non-BCa) percentile intervals are declined to be used for hypothesis testing. The aim is to consider the potential asymmetry of coefficients distributions. To do so, we first compute the one-sided p-value α^* that makes $\mathbf{c}[\alpha^*]_-$ equalize the value tested under the null, $\tilde{\theta}$. Formally,

$$\alpha^* = \underset{\alpha \in [0,1]}{\text{argeq}} \left(\mathbf{c}[\alpha^*]_- = \tilde{\theta} \right) = \Pr(\theta < \tilde{\theta})$$

We then compute $2 \min(\alpha^*, 1 - \alpha^*)$ to get the two-sided p-value associated to the bilateral version of the test.^d

^aWhere \mathbf{se} and \mathbf{c} are denoted by bold lowercase symbols to signal that they have the same dimension as their (column) vector input, with which they have a direct component-to-component mapping.

^bWhere, as for \mathbf{se} and \mathbf{c} , \mathbf{z} and \mathbf{a} are denoted by bold lowercase symbols to signal that they have the same dimension as their (column) vector input, with which they have a direct component-to-component mapping.

^cIndeed, if $\mathcal{I}_\gamma \cup \mathcal{I}_\rho \cup \mathcal{I}_\lambda = \emptyset$, eq.(2.28) boils down to eq.(2.23).

^dAn interesting discussion about taking twice the minimum as it stands can be found at <https://stats.stackexchange.com/a/140517>. As it reads there, the main justification of this approach is about complying with the fact that *cumulative distribution functions are invariant to order-preserving transformations*.

2.3.4 An application to the metropolitan area of Paris using the specification oknn

First, recall that in the scope of the econometric study *only*, one denotes the number of place by n instead of N_j as in the rest of the thesis.⁵³ Then, recall the expression of the indirect utility function in its private-vehicle specified form, *i.e.* considering eq.(1.102) with $u_{j,i}^0 = u_j^0$, that is

$$u_{j,i} = u_j^{0,2} R_{j,i}^{-\delta^H} \left(\frac{T}{d_{j,i}} v_{j,i} \right)^{\delta^a} (Y_j - P_{j,i}^a(a_{j,i}))^{1-\delta^a}$$

where $u_j^{0,2} = u_j^0 u_j^2$. Assuming that the metropolitan area of Paris is at equilibrium, *i.e.* $u_{\text{Paris},i} = u_{\text{Paris}} \forall i \in \llbracket 1; N_{\text{Paris}} \rrbracket$, and rearranging the expression to obtain housing rents per square meter on the LHS returns – one takes $j = \text{Paris}$ and omits the index in what follows for the sake of clarity,

$$\ln R_i = \beta_0 + \beta_1 \ln(Y - P_i^a) + \beta_2 \ln \left(\frac{T}{t_i} \right) + \varepsilon_i \quad (2.30)$$

with $\delta^H = 1/(\beta_1 + \beta_2)$ and $\delta^a = \beta_2/(\beta_1 + \beta_2)$.

Georeferenced Data

In Figure 2.6, we are shown that the number of places that are considered for the econometric study, 300, is not the one involved in deterministic simulations, 428. The reason of this

⁵³In the major part of this thesis, n_j and n respectively stand for the number of firms in agglomeration j and in the country of interest.

difference is two-fold: first, it is related to the different position-filters that have been used. Second, it is due to computation time concerns.

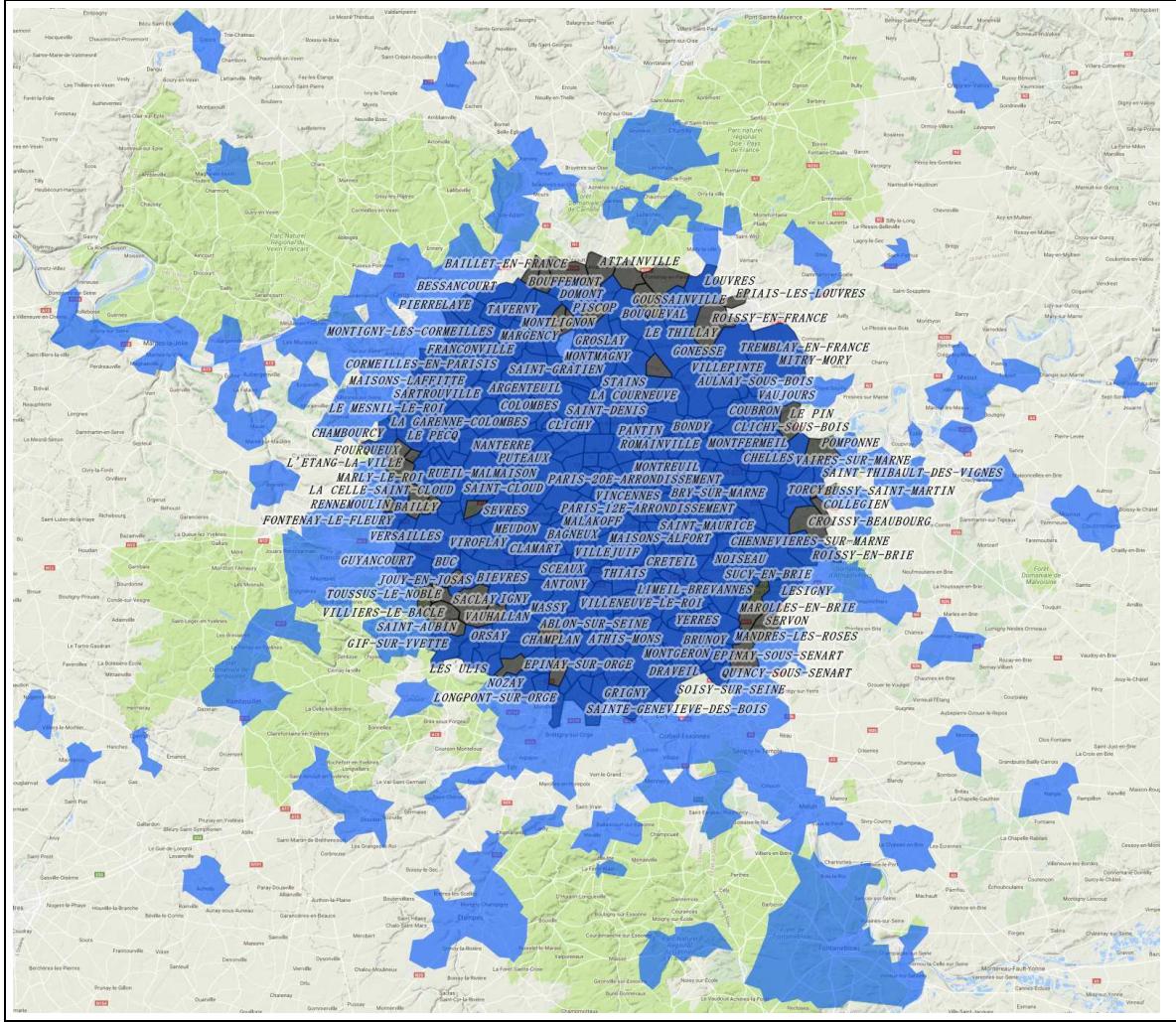


Figure 2.6: The 428 places considered in the simulation (blue) *versus* the 300 ones considered in the econometric study.

The 428 places that are considered during the simulation have been subject to the double filter of income availability on the one hand and of 95% population representativity on the other hand, which led to the emergence of geographic islands. On the contrary, the places that are considered in the econometric study are simply the 300 nearest ones to Place de l'Hôtel de ville.⁵⁴

The bootstrap resampling process in itself takes time, which is even more true when implementing Lin et al. (2007)'s, whose "main disadvantage is the high computing cost of large matrix inversion". Regardless of the chosen number of resamplings, SARIMA models are intrinsically costly to compute. Putting aside the cases of models based on long-distance autoregression

⁵⁴Wich may have caused the emergence of islands in other spaces, *e.g.* the plane formed by the two dimensions of populations and income.

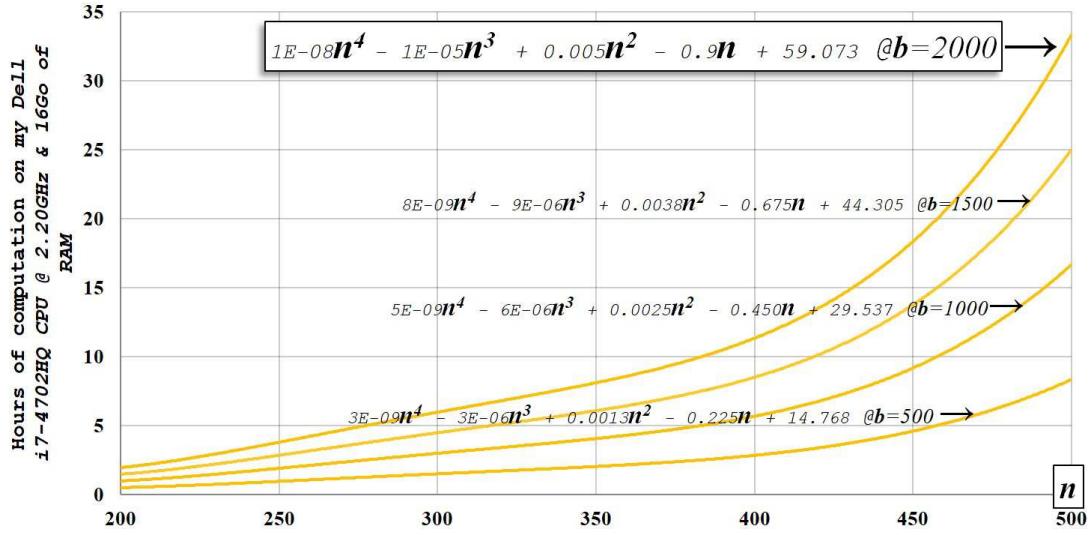


Figure 2.7: Polynomial Times of n

processes, *i.e.* SARIMA($\mathcal{I}_\rho, \emptyset, \emptyset$) and/or SARIMA($\emptyset, \mathcal{I}_\gamma, \emptyset$),⁵⁵ – that involves no $n \times n$ matrix inversions –, full SARIMA specifications analogous to that that are performed in TS are costly in terms of computation because of the MA terms. Indeed the consideration of the MA coefficients cannot be decoupled from the inversion of their framing $n \times n$ spatial filter, $\Lambda(\lambda)$, that is performed as many times as needed by the maximization of the log-likelihood. In place of the simplex search algorithm (Nelder and Mead, 1965) chosen for the maximization, techniques such as LeSage (1999, p.59-60)'s lattice logarithmic search could be envisaged when dealing with little sets of parameters, but has the disadvantage of becoming difficult to handle as the sets of parameters increase in size.⁵⁶ To finally justify the choice of 300 for the sample size, Figure 2.7 illustrates how costly the estimation procedure of a SARIMA({1}, {1}, {1}) is. We are shown that the time complexities of the input-size, n , and of the number of bootstrap samples, b , respectively are $O(n^4)$ and $O(b)$. b is set to 2000 in this work.

On these bases, we are shown in Figure 2.8 the two geographies of monthly housing rents and travel times, respectively bot-scraped from *LaCoteImmo*⁵⁷ and *GoogleMaps*. As can be easily seen from these two maps, housing rents and travel times both possess an auto-correlation of the same nature, *i.e.* following a long-distance process with a positive sign. Considered together under the (credible) assumption of no spurious relation, the two scatter graphs of these variables clearly illustrate the housing/transport trade-off given the opposite sign of their (road-distance supported) trends. The geography of rents - monotonously decreasing from the

⁵⁵Note that in the case in which the autoregression is in the endogenous variable, Anselin (1988) suggests some additional tricks to minimize at most the number of objects to compute when maximizing the log-likelihood.

⁵⁶First, note that the denomination that LeSage (1999) uses for his grid search is "multiple pass grid search" and second, it is primarily designed to avoid the repeated Jacobian evaluation involved when brute force maximizing the log-likelihood, be it SAR- or SEM- related thus. By citing this, we simply assume that such a techniques could also have been used to minimize the number of $n \times n$ matrix inversions needed when maximizing the log-likelihood of SMA models.

⁵⁷LaCoteImmo website is available at, *e.g.* <http://www.lacoteimmo.com/prix-de-l-immo/location/pays/france.htm>.

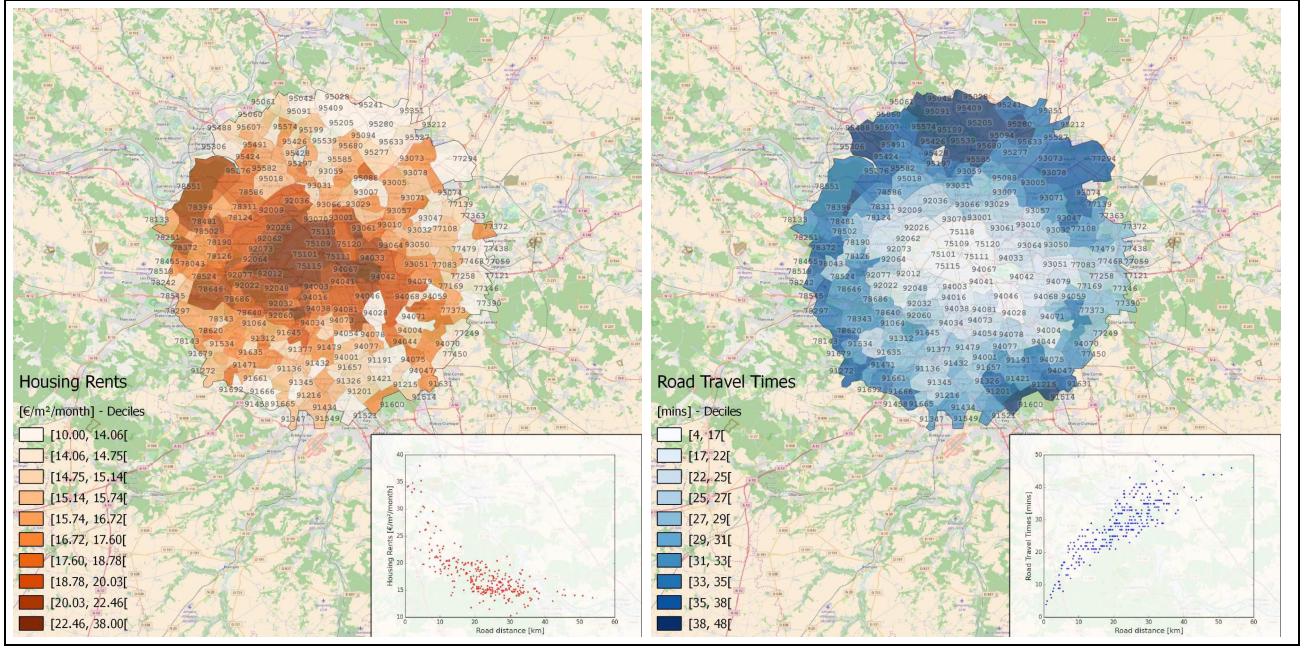


Figure 2.8: Housing rents, R_i , and Road travel times, t_i

center to the periphery – confirms that the choice of representing the metropolitan area of Paris with a monocentric LUTI-based spatial structure⁵⁸ is non-abusive, which suggests that the existence of a unique utility level of (urban) equilibrium covering the whole area is empirically valid.

In Figure 2.9, one is shown the two geographies of incomes net of transport cost and accessibility. Income, Y , derives on average⁵⁹ from income net of personal tax available at the level of municipalities in 2011, and is taken from *Direction Générale des Finances Publiques (DGFiP)*'s database.⁶⁰ In addition, note that Y is the result of a readjustment performed on incomes net of personal tax, originally calculated per fiscal household.⁶¹ P_i^a , computed as shown in eq.(1.25), is composed of a distance-variable part, that is $c_a p_a 251 d_i T / t_i$ where d_i is the road distance between place i and Place de l'Hôtel de ville and 251 stands for the number of business days used (and assumed constant over the prospective horizon) in GEMSE. c_a and p_a are set to their 2011 value, respectively 0.0678liter/km and 1.518€/liter. Table 2.1 reports the summary statistics of the variables implied in the study, sometimes implicitly as the private vehicle speeds are. This concludes the presentation of data used in the econometric study.

⁵⁸Recalling that the only numerical/mathematical objects that embody the spatial structure are the two geographies of rents and travel times, hence the expression "LUTI-based spatial structure".

⁵⁹Indeed, the average is considered since GEMSE does not space-differentiate incomes within a given urban area, hence $Y_i = Y \forall i = 1, \dots, N_{\text{Paris}}$.

⁶⁰Available at https://gemse.alwaysdata.net/static/gemse/FRANCE/impot_rev_2011.xlsx.

⁶¹In INSEE's 2011-data, income net of personal tax is denoted by RNNTFF11, while the number of households is denoted by NBFF11. This adjustment is also performed so as to comply with the nature of agents that are described at the urban level in GEMSE. Also recall that incomes that are involved in GEMSE's calibration are not directly taken from any database, but come from the spatial disaggregation of the GDP [see eq.(2.1) and eq.(2.4)].

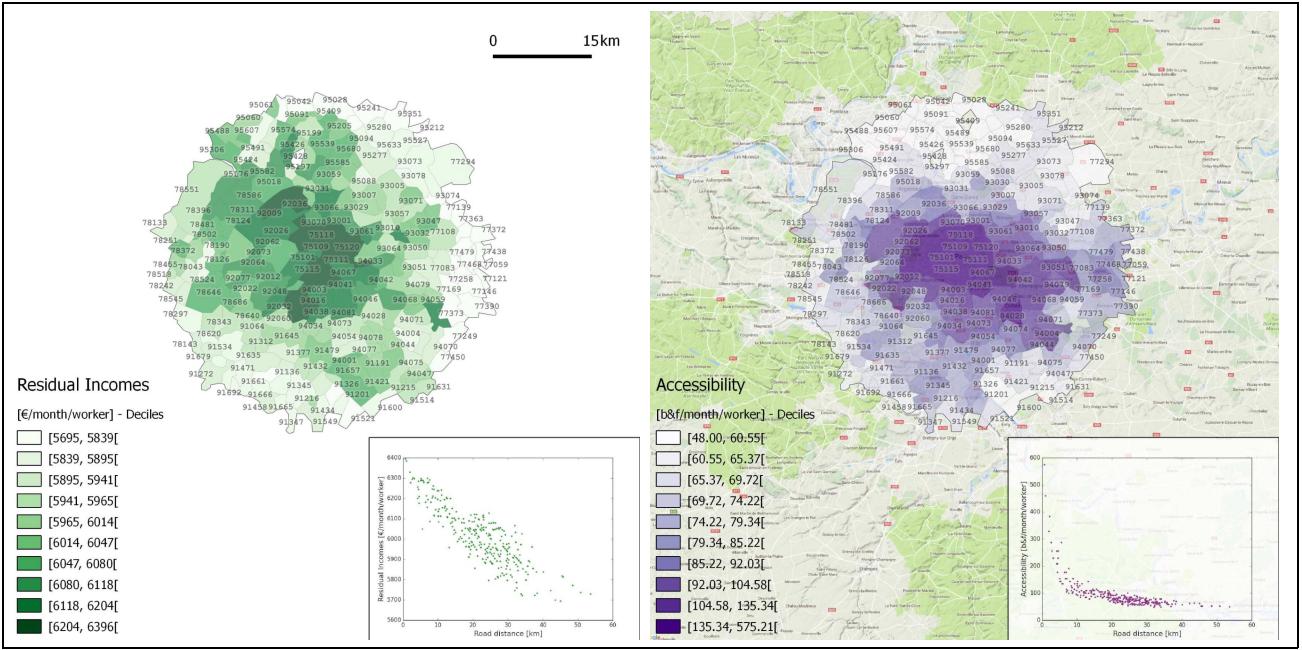


Figure 2.9: Residual incomes, $Y - P_i^a$, and Accessibility, T/t_i

Table 2.1: Summary statistics

$i = 1, \dots, 300$	Mean	St. dev.	Min	Max
R_i [€/m ² /year]	213.49	52.87	123.96	455.94
d_i [km]	22.10	9.72	0.70	53.60
t_i [mins]	28.51	8.34	4.00	48.00
v_i [km/h]	44.59	12.43	10.50	72.97
P_i^a [€/year]	6000.00	1673.26	1412.89	9819.28
$Y - P_i^a$ [€/year]	72160.44	1673.26	68341.16	76747.56
$251T/t_i$ [b&f/year]	1123.08	669.44	575.21	6902.50

SARIMA($\mathcal{I}_\rho, \mathcal{I}_\gamma, \mathcal{I}_\lambda$) model (i) Identification

We initiate the model identification with Table 2.2 that presents a first round estimation of eq.(2.30), with neither spatial components nor modeled residuals autocorrelation.⁶² Let's figure out whether there is a need to explicit the structure of autoregression or/and autocorrelation.

In Figure 2.10, one is shown four maps of SARIMA($\emptyset, \emptyset, \emptyset$)'s residuals for different definition of quantiles, that are, from the upper left to the lower right corner, biciles, terciles, quartiles and deciles. From this observation, one sees that the autocorrelation of residuals is very likely

⁶²Qualified as unmodeled, or equivalently, as modeled with an $n \times n$ identity matrix in lieu of correlation structure. Recall how important the assumptions of homoscedastic and unautocorrelated residuals are, since coefficients standard errors (and significance tests) directly inherit their reliability from the non-rejection of those. Indeed, in the general case (and \mathbf{X} considered fixed), recall that $\text{se}[\hat{\beta}]^2 = \text{diag}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]$ where $\hat{\Omega}$ is the (presumably misspecified) square correlation structure of residuals.

Table 2.2: SARIMA $(\emptyset, \emptyset, \emptyset)$

	Coeff.	Dist. ^a	Std. Err.	P> z	[95% C.I.]
β_0	-37.5566	D1	3.919830	0.000	-45.2708 -29.8425
		D2	.	.	.
		D3	3.952075	0.000	-45.5031 -29.6167
		D4	4.049645	0.000	-45.6052 -29.7309
β_1	3.6404	D1	0.357978	0.000	2.9359 4.3449
		D2	.	.	.
		D3	0.360783	0.000	2.9101 4.3689
		D4	0.368348	0.000	2.9339 4.3778
β_2	0.3133	D1	0.022422	0.000	0.2692 0.3574
		D2	.	.	.
		D3	0.022358	0.000	0.2692 0.3562
		D4	0.021910	0.000	0.2684 0.3542
δ^H	0.2529	D1	.	.	.
		D2	0.022155	0.000	0.2093 0.2965
		D3	0.022840	0.000	0.2148 0.3073
		D4	0.023482	0.000	0.2143 0.3064
δ^a	0.0792	D1	.	.	.
		D2	0.010986	0.000	0.0576 0.1009
		D3	0.011251	0.000	0.0602 0.1057
		D4	0.011360	0.000	0.0602 0.1047
σ_r^2	0.0146	D1	0.001189	.	.
		D2	.	.	.
		D3	0.001342	.	0.0118 0.0171
		D4	0.001350	.	0.0120 0.0173

^a C.I.-related objects derive from four types of distribution:

*D1 stands for normal-approximation-based distribution.

*D2 is as D1, but derives from an information matrix expressed in terms of δ^H and δ^a .

*D3 stands for bootstrap-based distribution. Associated std. are computed over bootstrap distributions.

*D4 stands for BCa bootstrap-based distribution. Associated std. are deduced by reversing the symmetry-based C.I. formula.

to be effective in the geographic space, and to follow a short distance process.

When observing biciles, which in this case is a manner to consider high orders of neighbors, it appears obvious that long-distance pairs of residuals do not follow a random process, but rather a negative autoregressive one. When considering terciles, which is a way of considering lower-than biciles related orders of neighbors, the spatial fragmentation occurs more among greater-than-0.002 (blue) biciles than among others (red), which suggests the existence of non-constant spatial variance that depends on the groups one forms. When considering quartiles,

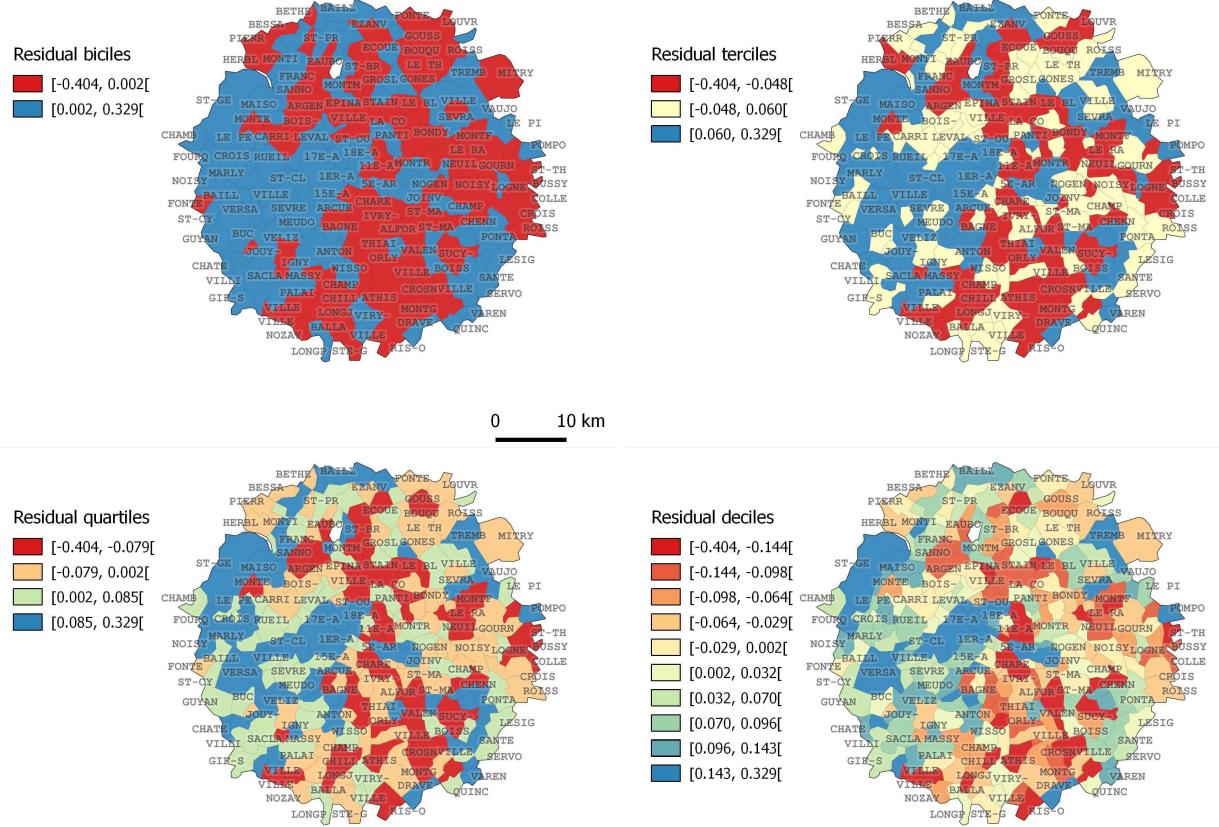


Figure 2.10: SARIMA($\emptyset, \emptyset, \emptyset$)-residual's mapped biciles, terciles, quartiles and deciles.

the most salient fact is that, while the blue space fragments, the red one does not change that much compared to the previous higher quantiles, *i.e.* rather than fragmenting, the red space, as a mountain range, undergoes a water-like surge that makes emerge islands. This suggests that this space concentrates rather outlying places. When observing deciles, one sees no long-distance gradient of colors, for example from the center to the periphery, which also enforces the assumption of the short-distance nature of the residual's autoregression. Note that this absence of long-distance trend is a notable difference with that of housing rents (see Map 2.8's left dial), residual incomes and accessibility (see Map 2.9's left and right dials) that do exhibit long-distance trends. This opposition strongly suggests that these three variables are cointegrated in space and that applying a SEM filter can improve the properties of the vector of residuals.

Undertaken within the stage of the identification of the differencing structure, Figure 2.11 shows SARIMA($\emptyset, \emptyset, \emptyset$)'s residuals plot and full autocorrelogram. Note the space-specific characteristics of the correlogram. It shows the so-called I of Moran (1950) and C of Geary (1954) that are associated to each neighbor-order, both transformed to be interterpretable as

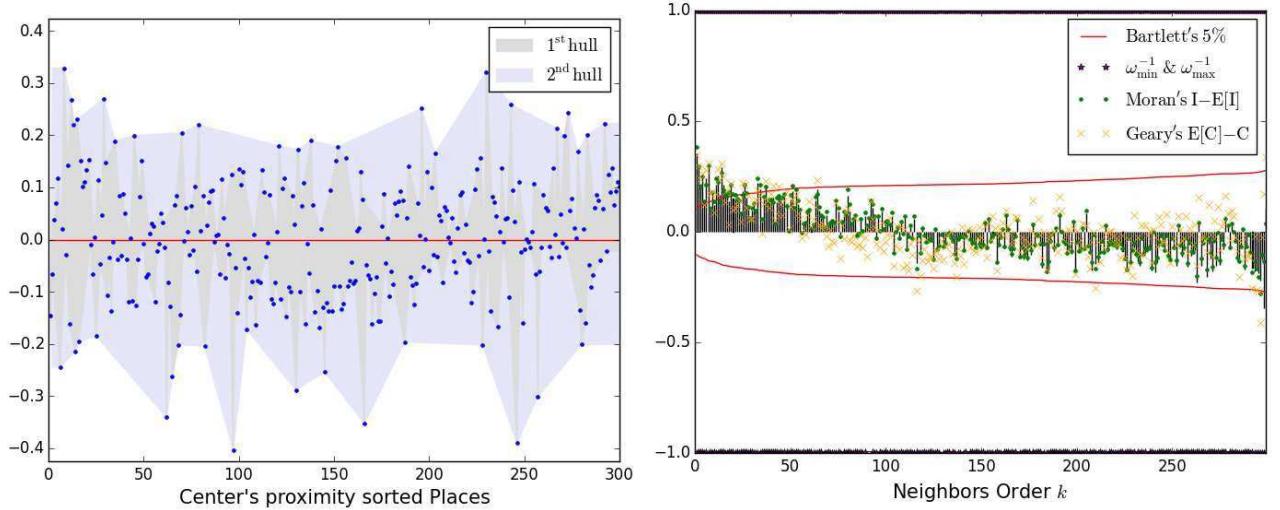


Figure 2.11: $\text{SARIMA}(\emptyset, \emptyset, \emptyset)$ residual's plot and (full) autocorrelogram

the regression coefficients are.⁶³ Looking closely at the definitions of I and C, one sees that the former – as a Pearson-like correlation coefficient – involves deviation from the sample mean while the latter involves deviations from neighbors values. Put differently, the former is more a global measure of autocorrelation than the latter.

The residuals plot confirms what was previously suggested regarding the (non-fragmenting) red space, *i.e.* this space contains places that instill heteroscedasticity. This is highlighted by the two hulls of the residuals,⁶⁴ especially the second one, in which one perceives the strong influence of some outlying negative deviations. The autocorrelogram in addition shows that this non-constant variance may also be accompanied by multiple local spatial trends, given the multiple slow linear decay patterns starting from, *e.g.* the 1st, 7th, 14th, 19th, 26th and 37th orders of neighbors.

Figures 2.12 and 2.13 zoom at the 50 first neighbors orders, and juxtapose two versions of correlogram versions whose difference lays in their abscissa. In these two figures, the (partial and full) Auto Correlation Function (ACF) plots on the left dial are – as in the TS case – based on neighbors orders while that on the right dial is based on average inter-distance neighbors.⁶⁵ This allows us to rephrase the interpretation of the ACF in terms of kilometers in the *one-dimensional space of inter-distances*, as follows: there exist multiple local spatial trends that start from, approximately, 1.8km, 3.8km, 5.5km, 6.5km, 7.7km and 9.3km. *What is*

⁶³Moran's I usually ranges from -1 to 1 but is *a fortiori* not 0-centered, hence $I - \mathbb{E}[I]$, where $\mathbb{E}[I] = -(n-1)^{-1}$ is the expected value of I under the null hypothesis of no spatial autocorrelation. Geary's C usually ranges from 0 to 2 . Values that are below (resp. above) 1 indicate positive (resp. negative) spatial autocorrelation. Thus, $\mathbb{E}[C] - C$ ranges from -1 to 1 and is 0-centered, where $\mathbb{E}[C] = 1$ is the expected value of C under the null hypothesis of no spatial autocorrelation.

⁶⁴The first hull is formed over the local extrema, while the second hull is formed over the local extrema of the first-hull zero-centered extrema.

⁶⁵Note that in the TS case, this nuance makes no sense since temporal positions are traditionally chosen to be equidistant.

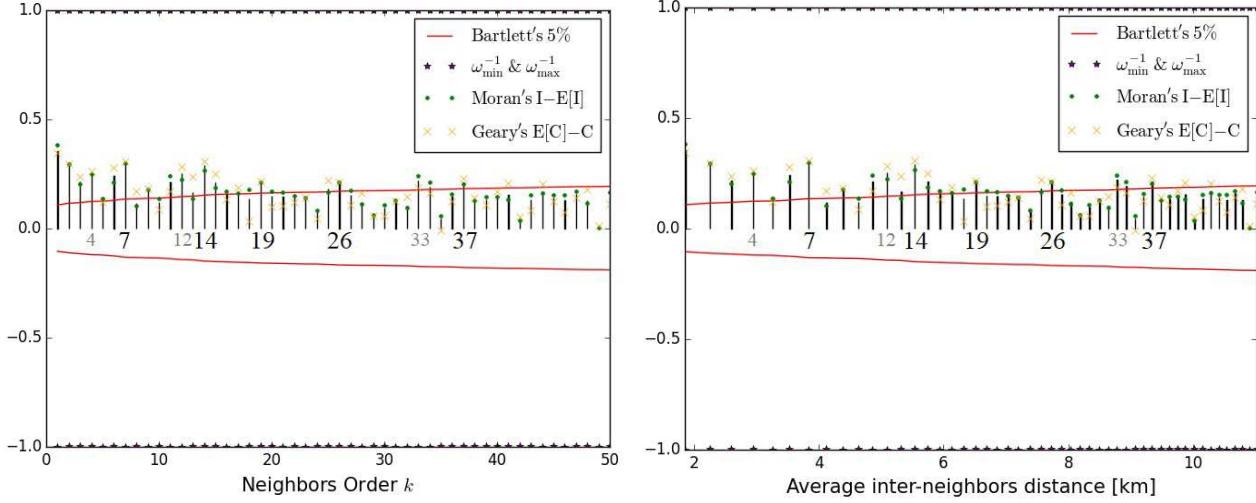


Figure 2.12: SARIMA $(\emptyset, \emptyset, \emptyset)$ residual's full autocorrelogram against orders (left) and kilometers (right)

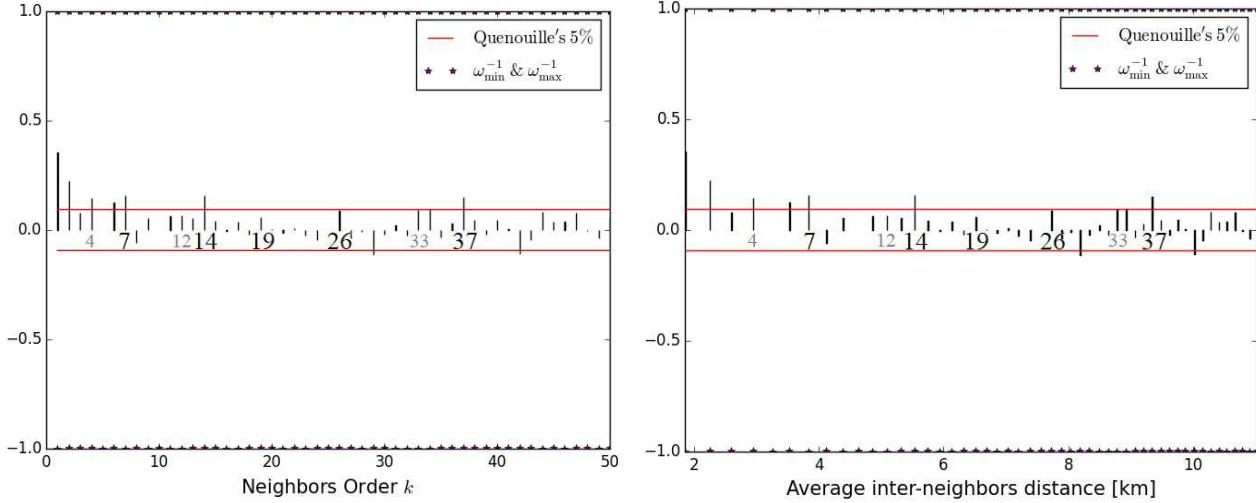


Figure 2.13: SARIMA $(\emptyset, \emptyset, \emptyset)$ residual's partial autocorrelogram against orders (left) and kilometers (right)

happening there? At order 1 or at km 1.8, wherever one is located in the agglomeration, one is positioned on the average centroid of the two closest neighbors. Interpreted in terms of travel times or housing rents, it appears reasonable to think that no clearly perceptible structural break exists (on average) at such a short distance. Thus, this first neighbor-order is likely to be representative of the reciprocal short distance effect between places. At order 7 or at km 3.8, if one takes the 1st arrondissement as a distance origin, one is just before the *north-south* frontier of intra-mural Paris. Still at order 7, but this time taking an example outside of intra-mural Paris, one is very likely to have moved to a new quartile (see Figure 2.10). This suggests that the 7th neighbor-order may coincide with a structural break both in housing rents and traffic conditions. The existence of such a structural break is also suggested when taking a look at the

geographies of housing rents (see Map 2.8's left dial), residual incomes and accessibility (see Map 2.9). Indeed, by randomly drawing a place and considering a 3.8km-radius circle around it, one is likely to get a group of places belonging to the same decile. At order 14 – which in the TS case would seasonally superimposes to the just-explained 7th order – or at km 5.5, one is very likely to capture the same sort of information as the one that is captured at order 7 and whose value-added stems from the irregularity of the lattice. Note that by taking again the 1th arrondissement as a distance origin, the 14th order delimits the *north-south outer-side* frontier of intra-mural Paris. Averaged together, the 7th and the 14th neighbors orders are, remarkably, equivalent to the km 4.6, which is the real *north-south* radius of intra-mural Paris. At order 19 (km 6.5) or 26 (km 7.7), the information captured is not that clear and we thus skip these neighbors order. At order 37 or at km 9.3, one is actually looking at the last significant order stated by the ACF plot. Remarkably, once again with the 1st arrondissement as a distance origin, know that 9.3km is the real *east-west* radius of intra-mural Paris.

Particularly keen not to resort to any kind of automatic selection techniques, it is on these geographically established fundamentals, that we define the structure of the 1st difference over lags 7, 14 and 37, *i.e.* $\mathcal{I}_\gamma = \{7, 14, 37\}$, all three representative of the north-south and east-west dimensions of intra-mural Paris. We justify our choice of defining \mathcal{I}_γ as it stands by the significance of these 3 lags both on the PACF and ACF sides. We then consider the 1st, 4th and 12th neighbor-orders as a starting point to structure the SMA part of the DGP, *i.e.* $\mathcal{I}_\lambda = \{1, 4, 12\}$.

SARIMA($\emptyset, \{7, 14, 37\}, \{1, 4, 12\}$)-grounded models parameters *(ii)* Estimation and *(iii)* Checking

Figure 2.14 shows the ACF and PACF resulting from the inclusion of these lags in the model. Based on the interpretation of these two correlograms, multiple possibilities arise. Table 2.3 shows some selective properties of the so-suggested models. Multiple model-specifications may reasonably well explain the generative process of housing rents, explicated in eq.(2.30). It thus appears informative to compare their results.

We do not compute the bootstrap intervals of all the models that are enumerated in Table 2.3, only the first three ranked ones. The ranking process is five-dimensional and performed according to the so-defined lexical order: SH-pv > 5%, BP-pv > 5%, BIC, SH-pv, BP-pv. For example, the 5d rank of the 1st model is *True, True, -443.18, 7.13%, 11.26%*. That of the 14th is *False, True, -444.91, 3.3%, 17.52%*. Note the importance given to the non-rejection of prediction errors' normality and residuals homoscedasticity. For the former non-rejection, this is justified by the Gaussian nature of the maximized (log) likelihood, on which information criteria rely. The importance given to the latter non-rejection is justified by the simple fact that

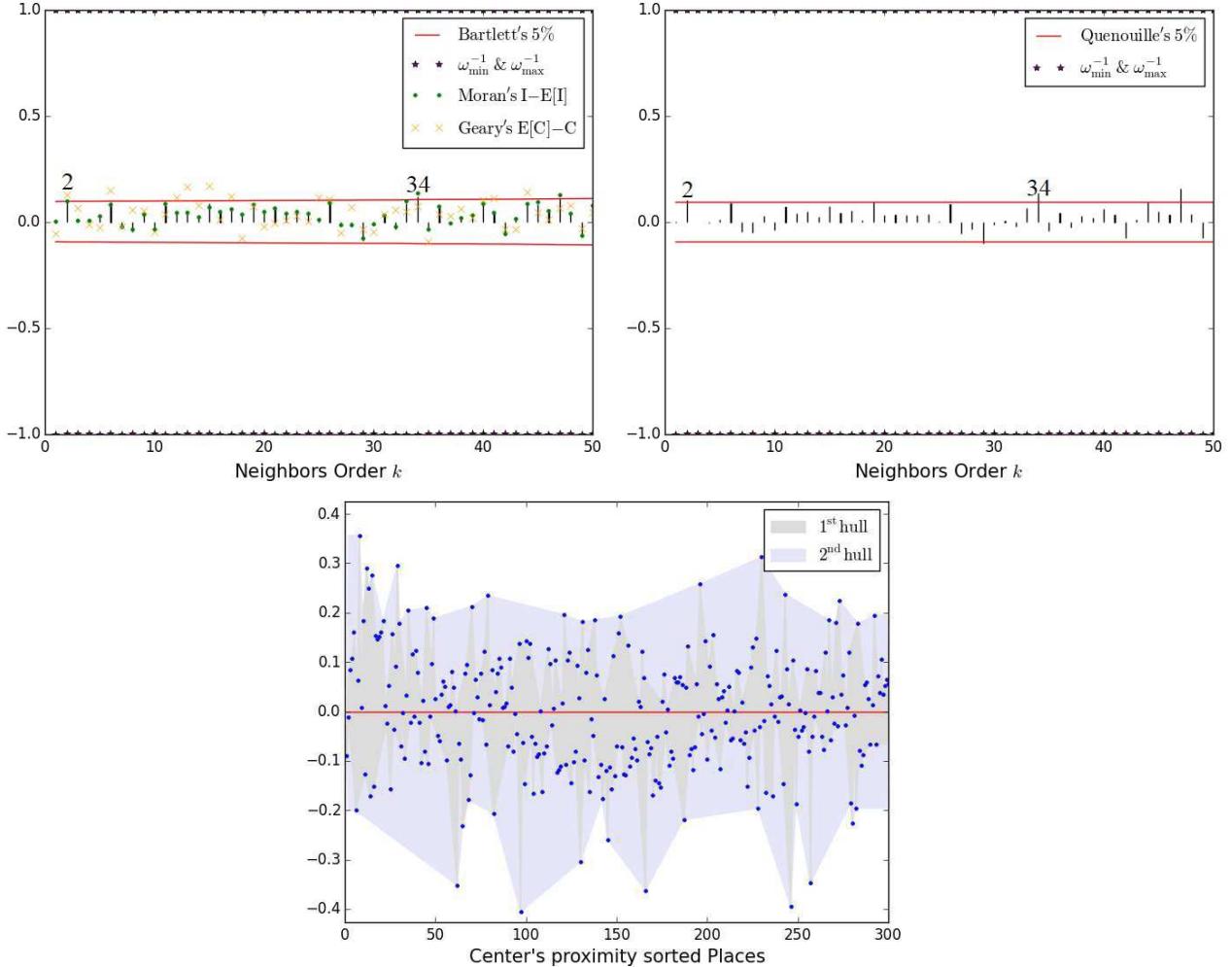


Figure 2.14: $\text{SARIMA}(\emptyset, \{7, 14, 37\}, \{1, 4, 12\})$ residual's full (left), partial autocorrelograms and residual plot

the reliability of the non-percentile-based standards errors heavily depends on it.⁶⁶ In what follows, we first check the properties of these three models, and we then consider especially the estimated values for δ^H and δ^a , whose bootstrap-based density plots are shown in Figure 2.15.

Relatively to model checking, we first refer to the (vertically rotated) left parts of Tables 2.4, 2.5 and 2.6. In those, one can read that all the unit-root tests lead to strongly reject the null according to which spatial coefficients are equal either to the lower or the upper bound of their stationary space. But it is not sufficient and one must ensure that all spatial filters are non-singular. To do so, we then consider the sum of the absolute values of the estimated coefficients belonging to the same family, *i.e.* involved in the same spatial filter. It appears that none of these sums approach 1.

Relatively to housing services, the estimated values for δ^H – all greater than that of 25% set in GEMSE for France –, are in accordance with a study realized by *Institut national de la*

⁶⁶As already mentioned in a previous footnote, in the general case (and \mathbf{X} considered fixed), recall that $\text{se}[\hat{\beta}]^2 = \text{diag}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]$ where $\hat{\Omega}$ is the (presumably misspecified) square correlation structure of residuals.

Table 2.3: SARIMA($\emptyset, \{7, 14, 37\}, \{1, 4, 12\}$)-grounded eligible models

SARIMA ($\mathcal{I}_\rho, \mathcal{I}_\gamma, \mathcal{I}_\lambda$)	(P)ACF score ^a	Rank	BIC	Selection criteria			Fit criteria		Residuals properties	
				AIC	HQC ^b	$\ln \mathcal{L}$	MSE	R ²	BP-pv ^c	SH-pv ^d
($\emptyset, \{2, 7, 14, 34, 37\}, \{1, 4, 12\}$)	62.97	1	-443.18	-483.93	-467.62	252.96	0.0114	0.6800	0.1126	0.0713
($\emptyset, \{7, 14, 37\}, \{1, 4, 12\}$)	64.35	2	-442.61	-475.94	-462.60	246.97	0.0118	0.6826	0.0902	0.0995
($\{2\}, \{7, 14, 37\}, \{1, 4, 12\}$)	62.23	3	-442.40	-479.44	-464.62	249.72	0.0115	0.6886	0.0897	0.0734
($\{2\}, \{7, 14, 34, 37\}, \{1, 4, 12\}$)	62.29	4	-441.75	-482.49	-466.19	252.25	0.0113	0.6895	0.1119	0.0569
($\emptyset, \{2, 7, 14, 37\}, \{1, 4, 12\}$)	64.38	5	-441.33	-478.37	-463.55	249.19	0.0116	0.6815	0.0970	0.0831
($\{2\}, \{7, 14, 37\}, \{1, 4, 12, 33\}$)	63.70	6	-440.08	-480.82	-464.51	251.41	0.0114	0.6881	0.0990	0.0654
($\{2\}, \{7, 14, 37\}, \{1, 4, 12, 34\}$)	59.41	7	-439.29	-480.03	-463.73	251.02	0.0114	0.6896	0.0843	0.0546
($\{33\}, \{7, 14, 37\}, \{1, 2, 4, 12\}$)	64.38	8	-438.21	-478.95	-462.64	250.47	0.0117	0.6924	0.0653	0.0584
($\{34\}, \{7, 14, 37\}, \{1, 2, 4, 12\}$)	60.14	9	-434.48	-475.22	-458.91	248.61	0.0117	0.6847	0.1355	0.0732
($\emptyset, \{7, 14, 37\}, \{1, 4, 12, 33\}$)	62.97	10	-431.75	-468.78	-453.96	244.39	0.0121	0.6841	0.0922	0.1149
($\{33\}, \{7, 14, 37\}, \{1, 4, 12\}$)	58.69	11	-435.76	-472.79	-457.97	246.40	0.0119	0.6904	0.0420	0.0676
($\emptyset, \{7, 14, 37\}, \{1, 2, 4, 12\}$)	59.41	12	-434.94	-471.98	-457.16	245.99	0.0123	0.6838	0.0298	0.1041
($\emptyset, \emptyset, \emptyset$)	48.83	13	-400.38	-411.49	-407.04	208.74	0.0147	0.6872	0.0367	0.1716
($\{2, 34\}, \{7, 14, 37\}, \{1, 4, 12\}$)	62.94	14	-444.91	-485.65	-469.35	253.83	0.0113	0.6861	0.1752	0.0330
($\{2, 33\}, \{7, 14, 37\}, \{1, 4, 12\}$)	62.97	15	-441.81	-482.55	-466.24	252.27	0.0113	0.6959	0.0799	0.0323

^a (P)ACF score *norm*-considers the number of non-significant lags in the ACF and in the PACF, $\#\text{lag}_{(P)\text{ACF}}$, via $(\#\text{lag}_{(P)\text{ACF}}^2 + \#\text{lag}_{\text{PACF}}^2)^{-\frac{1}{2}}$.

^b HQC stands for the Hannan-Quinn information criterion (Hannan and Quinn, 1979).

^c BP-pv stands for the p-value associated to Breusch and Pagan (1979)'s test statistic whose null hypothesis is *model residuals'* homoskedasticity.

^d SH-pv stands for the p-value associated to Shapiro and Wilk (1965)'s test statistic whose null hypothesis is *prediction errors'* normality.

statistique et des études économiques (INSEE) (Durand, 2012, p.9) with 2008-data. As it reads in there, "la part du loyer moyen dans le revenu des ménages interrogés s'élève à 34%". When referring to the (vertically rotated) right parts of Tables 2.4, 2.5 and 2.6, it appears that the possibility of having 25% as housing budget share is always rejected at a 5% significance level. However, it appears reasonable to assume that expanding the number of places from 300 to the one set in GEMSE, 428, would exert an upward pressure on these p-values, since the places that would be so-considered are very likely to exhibit economic characteristics that would in turn downward-pressure Paris's averages (of these economic characteristics) toward France's. Relatively to transport services, the level of 10% that is set in GEMSE is remarkably close to those yielded by the three specifications, all highly significant.

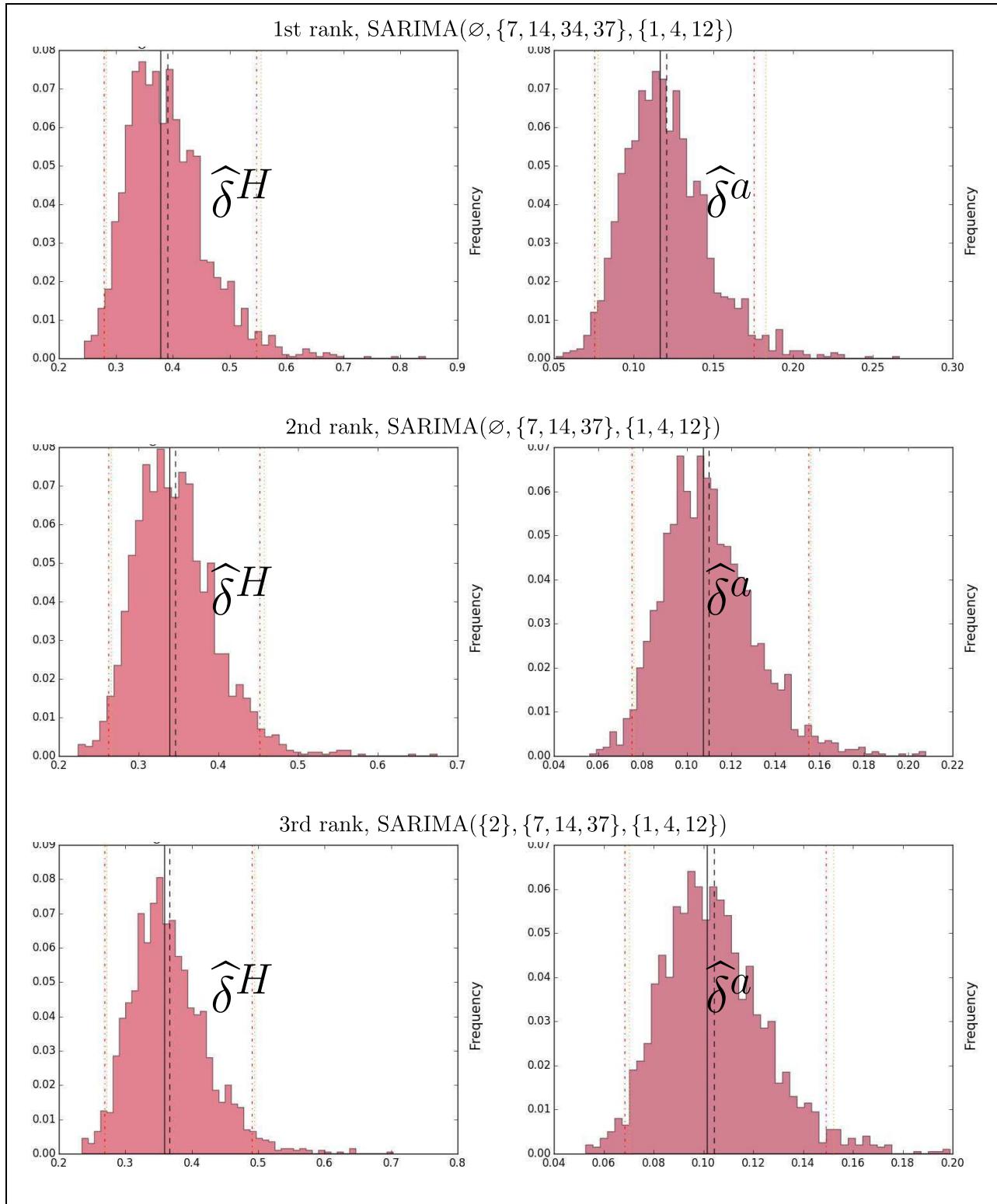


Figure 2.15: Bootstrapped budget shares' density plots

2.3.5 Main findings

Three mains findings are presented to conclude this section. The first is related to the simplicity of the structural equation that is estimated. The second relates to the estimated values themselves, similar to what one finds in the literature. The third finding relates to the proposal of lag operator that is implemented to conduct the validation of the budgets shares that are set in GEMSE.

The equation used in the study is really simple since only bivariate. It derives from the (indirect) utility function employed in GEMSE, rearranged to express housing rents [see eq.(2.30)]. This is a notable difference with the technique that is traditionally employed when dealing with housing rents or prices, such as hedonic regressions that rely on a (very) large set of characteristics, rarely publicly or easily available.

This structural equation remarkably yields estimated values for parameters that are similar to what one finds in the literature regarding France and Paris in particular. Indeed, the values found in the literature for France [see *e.g.* Guidetti and INSEE (2012, p.3, Table 3) or Morer and INSEE (2015, p.3, Table 4)] (and set in GEMSE) of 10% and 25% respectively for δ^a and δ^H are statistically supported in this study on the transportation side, and not so far from being so on the housing side. For Paris, which differs from the rest of the country on the housing side, our study remarkably finds (confidence and percentile) intervals that contain almost into their center the values found in the literature of 34% [see (Durand, 2012, p.9)].

On the theoretical plan, our econometric study generalizes in space an approach traditionally used in time series and presents the strict spatial counterpart of the well-known time-lag operator, baptized *k-nearest neighbor only (oknn)*. Remarkably, it allows us to identify some key elements of the spatial structure of Paris, namely its North-South and East-West intra-mural dimensions. Given that the approach that we develop requires careful handling, it is only applied to this urban area. Its (automation if feasible and) declination into spatial panel is a promising area for future research.

2.A Models selected for estimation

The models that have been chosen for estimation on the basis of their characteristics [see Table 2.3] are presented sequentially in what follows.

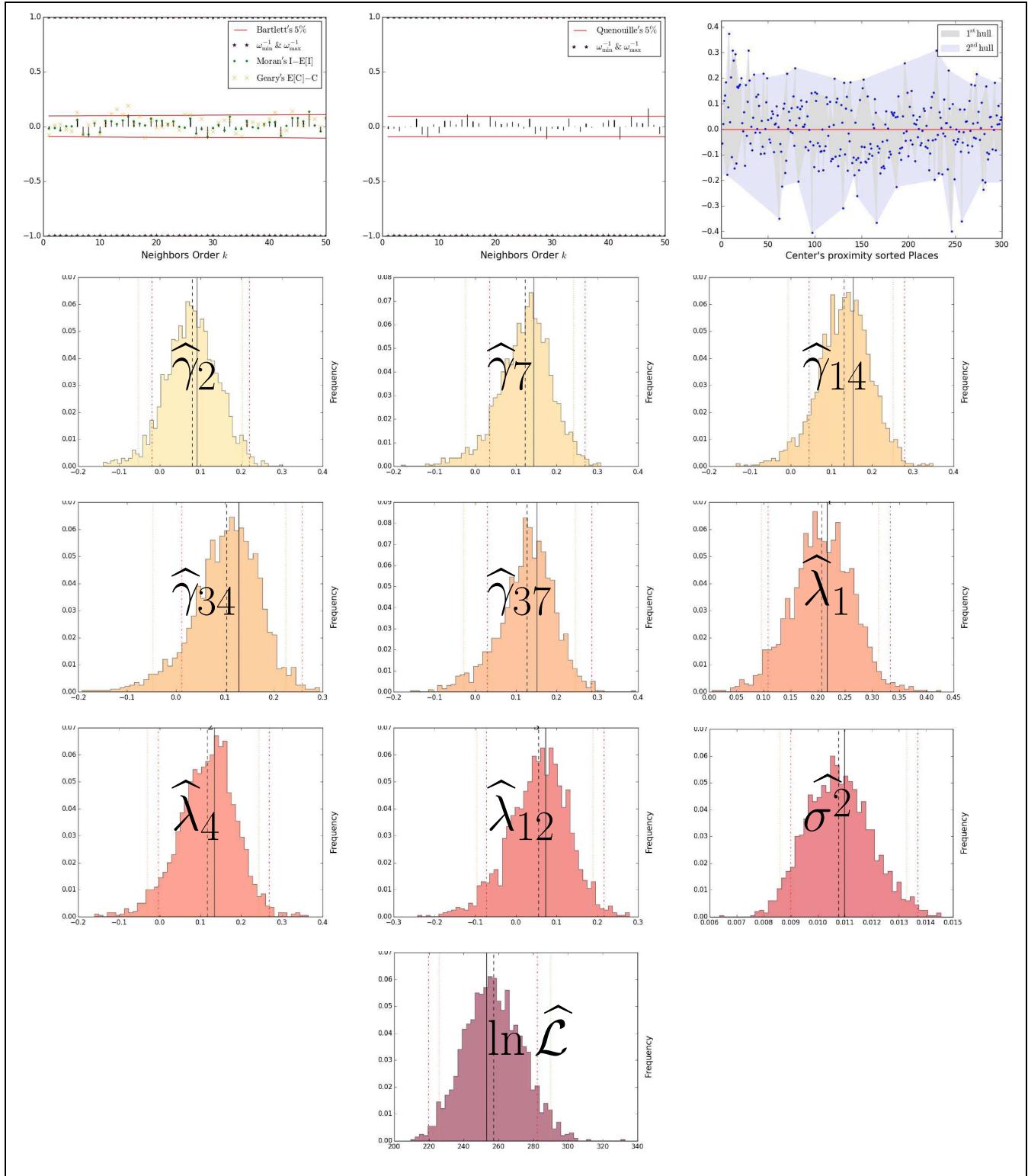


Figure 2.16: SARIMA($\emptyset, \{7, 14, 34, 37\}, \{1, 4, 12\}$)'s (P)ACFs, residual plot and bootstrap-based density plots of spatial parameters (1st rank)

Table 2.4: SARIMA ($\emptyset, \{2, 7, 14, 34, 37\}, \{1, 4, 12\}$)

	Coeff.	Dist. ^a	Std. Err.	P> z	[95% C.I.]
β_0	-22.9356	D1	4.677358	0.000	-32.1416 -13.7295
		D2	.	.	.
		D3	4.984052	0.000	-32.9913 -13.4897
		D4	5.022955	0.000	-33.5484 -13.8588
β_1	2.3373	D1	0.420121	0.000	1.5104 3.1641
		D2	.	.	.
		D3	0.448576	0.000	1.4777 3.2423
		D4	0.452586	0.000	1.5387 3.3128
β_2	0.3085	D1	0.033398	0.000	0.2427 0.3742
		D2	.	.	.
		D3	0.034433	0.000	0.2401 0.3761
		D4	0.034243	0.000	0.2396 0.3738
γ_2	0.0914	D1	0.050415	0.070	-0.0078 0.1906
		D2	.	.	.
		D3	0.063333	0.185	-0.0533 0.2027
		D4	0.060998	0.088	-0.0195 0.2196
γ_7	0.1431	D1	0.049822	0.004	0.0451 0.2412
		D2	.	.	.
		D3	0.064895	0.081	-0.0238 0.2410
		D4	0.059817	0.019	0.0351 0.2696
γ_{14}	0.1531	D1	0.050293	0.002	0.0541 0.2521
		D2	.	.	.
		D3	0.064489	0.064	-0.0065 0.2510
		D4	0.059783	0.014	0.0446 0.2789
γ_{34}	0.1281	D1	0.046504	0.006	0.0366 0.2197
		D2	.	.	.
		D3	0.067534	0.145	-0.0488 0.2242
		D4	0.062970	0.039	0.0111 0.2579
γ_{37}	0.1513	D1	0.049874	0.002	0.0532 0.2495
		D2	.	.	.
		D3	0.067472	0.094	-0.0286 0.2453
		D4	0.065531	0.024	0.0291 0.2860
λ_1	0.2169	D1	0.048663	0.000	0.1211 0.3127
		D2	.	.	.
		D3	0.056515	0.000	0.0960 0.3110
		D4	0.057333	0.000	0.1083 0.3331
λ_4	0.1343	D1	0.052664	0.011	0.0306 0.2379
		D2	.	.	.
		D3	0.070150	0.114	-0.0311 0.2437
		D4	0.069426	0.254	-0.0037 0.2685
λ_{12}	0.0722	D1	0.052539	0.170	-0.0312 0.1756
		D2	.	.	.
		D3	0.073198	0.427	-0.0959 0.1883
		D4	0.073365	0.247	-0.0722 0.2154
δ^H	0.3780	D1	.	.	.
		D2	0.059715	0.000	0.2604 0.4955
		D3	0.070674	0.000	0.2819 0.5541
		D4	0.068306	0.000	0.2793 0.5470
δ^a	0.1166	D1	.	.	.
		D2	0.022648	0.000	0.0720 0.1612
		D3	0.026235	0.000	0.0776 0.1826
		D4	0.025519	0.000	0.0755 0.1756
σ_r^2	0.0110	D1	0.000913	.	.
		D2	.	.	.
		D3	0.001191	.	0.0086 100.0133
		D4	0.001202	.	0.0090 0.0137

^a C.I.-related objects derive from four types of distribution:

*D1 stands for normal-approximation-based distribution.

*D2 is as D1, but derives from an information matrix expressed in terms of δ^H and δ^a .

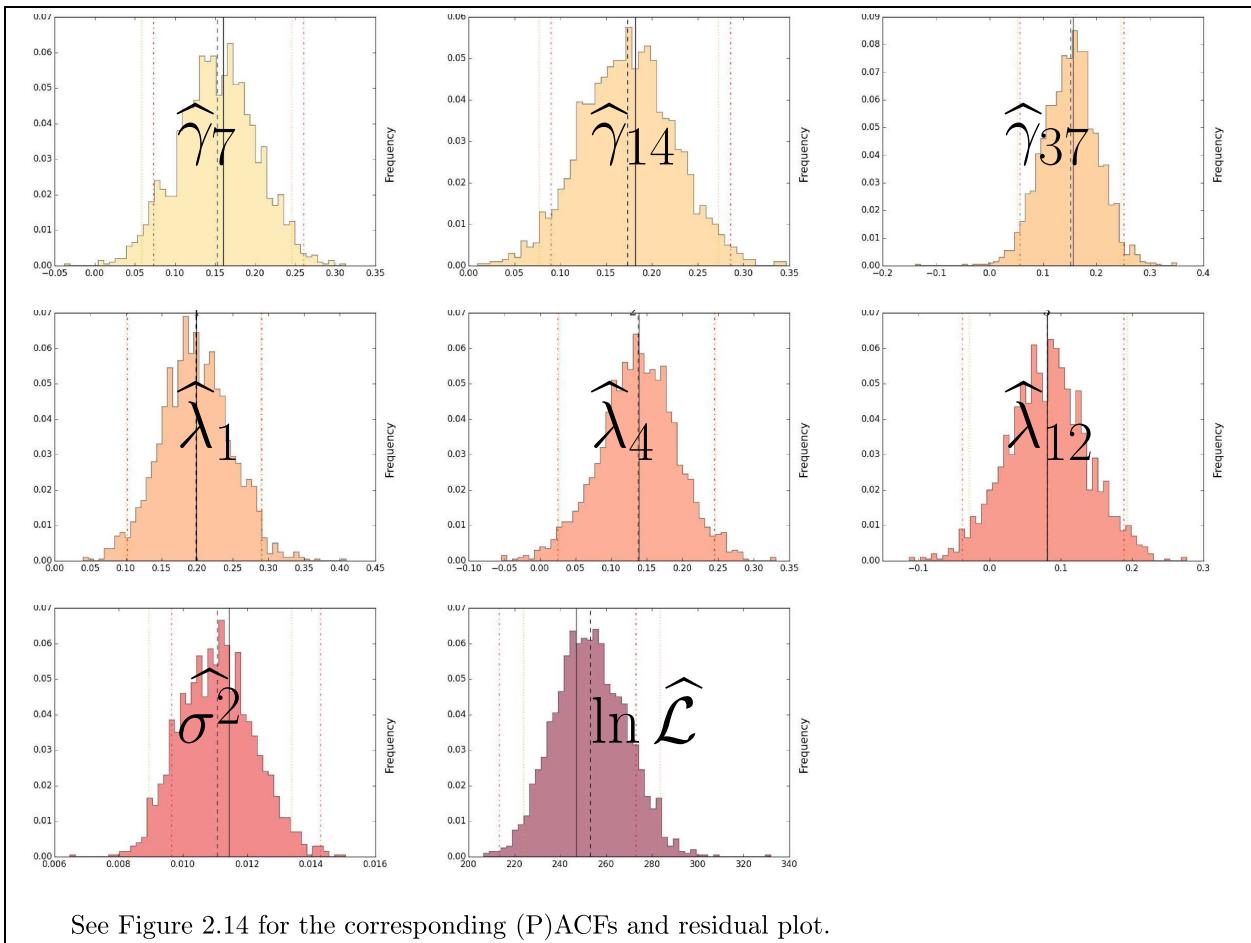


Figure 2.17: SARIMA($\emptyset, \{7, 14, 37\}, \{1, 4, 12\}$)'s (P)ACFs, residual plot and bootstrap-based density plots of spatial parameters (2nd rank)

Table 2.5: SARIMA $(\emptyset, \{7, 14, 37\}, \{1, 4, 12\})$

Coeff.	Dist. ^a	Std. Err.	P> z	[95% C.I.]	
β_0	D1	4.876331	0.000	-35.9117	-16.7170
	D2
	D3	4.627975	0.000	-35.4831	-17.7473
	D4	4.617045	0.000	-36.1738	-18.0754
β_1	D1	0.439771	0.000	1.7686	3.4997
	D2
	D3	0.417639	0.000	1.8658	3.4698
	D4	0.412382	0.000	1.8936	3.5101
β_2	D1	0.030931	0.000	0.2557	0.3775
	D2
	D3	0.030925	0.000	0.2584	0.3776
	D4	0.030529	0.000	0.2574	0.3771
γ_7	D1	0.050056	0.001	0.0616	0.2586
	D2
	D3	0.048355	0.001	0.0586	0.2452
	D4	0.047785	0.000	0.0733	0.2606
γ_{14}	D1	0.052825	0.001	0.0777	0.2856
	D2
	D3	0.050199	0.000	0.0767	0.2725
	D4	0.050028	0.000	0.0897	0.2858
γ_{37}	D1	0.051381	0.002	0.0549	0.2572
	D2
	D3	0.051040	0.005	0.0519	0.2449
	D4	0.049602	0.005	0.0566	0.2511
λ_1	D1	0.045590	0.000	0.1090	0.2885
	D2
	D3	0.048764	0.000	0.0999	0.2892
	D4	0.048213	0.000	0.1014	0.2904
λ_4	D1	0.051791	0.007	0.0366	0.2405
	D2
	D3	0.054354	0.016	0.0263	0.2464
	D4	0.056146	0.017	0.0249	0.2449
λ_{12}	D1	0.052018	0.119	-0.0213	0.1834
	D2
	D3	0.056203	0.150	-0.0290	0.1929
	D4	0.057675	0.170	-0.0380	0.1881
δ^H	D1
	D2	0.049880	0.000	0.2407	0.4371
	D3	0.050861	0.000	0.2650	0.4574
	D4	0.048467	0.000	0.2618	0.4518
δ^a	D1
	D2	0.020214	0.000	0.0675	0.1471
	D3	0.020579	0.000	0.0759	0.1558
	D4	0.020385	0.000	0.0750	0.1550
σ_r^2	D1	0.000944	.	.	.
	D2
	D3	0.001151	.	0.0089	0.0134
	D4	0.001188	.	0.0096	0.0143

Coeff.	Dist. ^a	P> z			
		$\mathcal{H}_0 : \hat{\gamma} = 0.05$	$\mathcal{H}_0 : \hat{\gamma} = 0.12$	$\mathcal{H}_0 : \hat{\gamma} = 0.15$	$\mathcal{H}_0 : \hat{\gamma} = 0.25$
γ_7	D3	0.0000	.	.	0.0000
γ_{14}	D3	0.0000	.	.	0.0000
γ_{37}	D3	0.0000	.	.	0.0000
λ_1	D3	0.0000	.	.	0.0000
λ_4	D3	0.0000	.	.	0.0000
λ_{12}	D3	0.0000	.	.	0.0000
δ^H	D3	0.0000	.	.	0.0000
δ^a	D3	0.0000	0.5634	.	0.0000
	D4	0.0000	0.5276	.	0.0000

^a C.I.-related objects derive from four types of distribution:

*D1 stands for normal-approximation-based distribution.

*D2 is as D1, but derives from an information matrix expressed in terms of δ^H and δ^a .

*D3 stands for bootstrap-based distribution. Associated std. are computed over bootstrap distributions.

*D4 stands for BCa bootstrap-based distribution. Associated std. are deduced by reversing the symmetry-based C.I. formula.

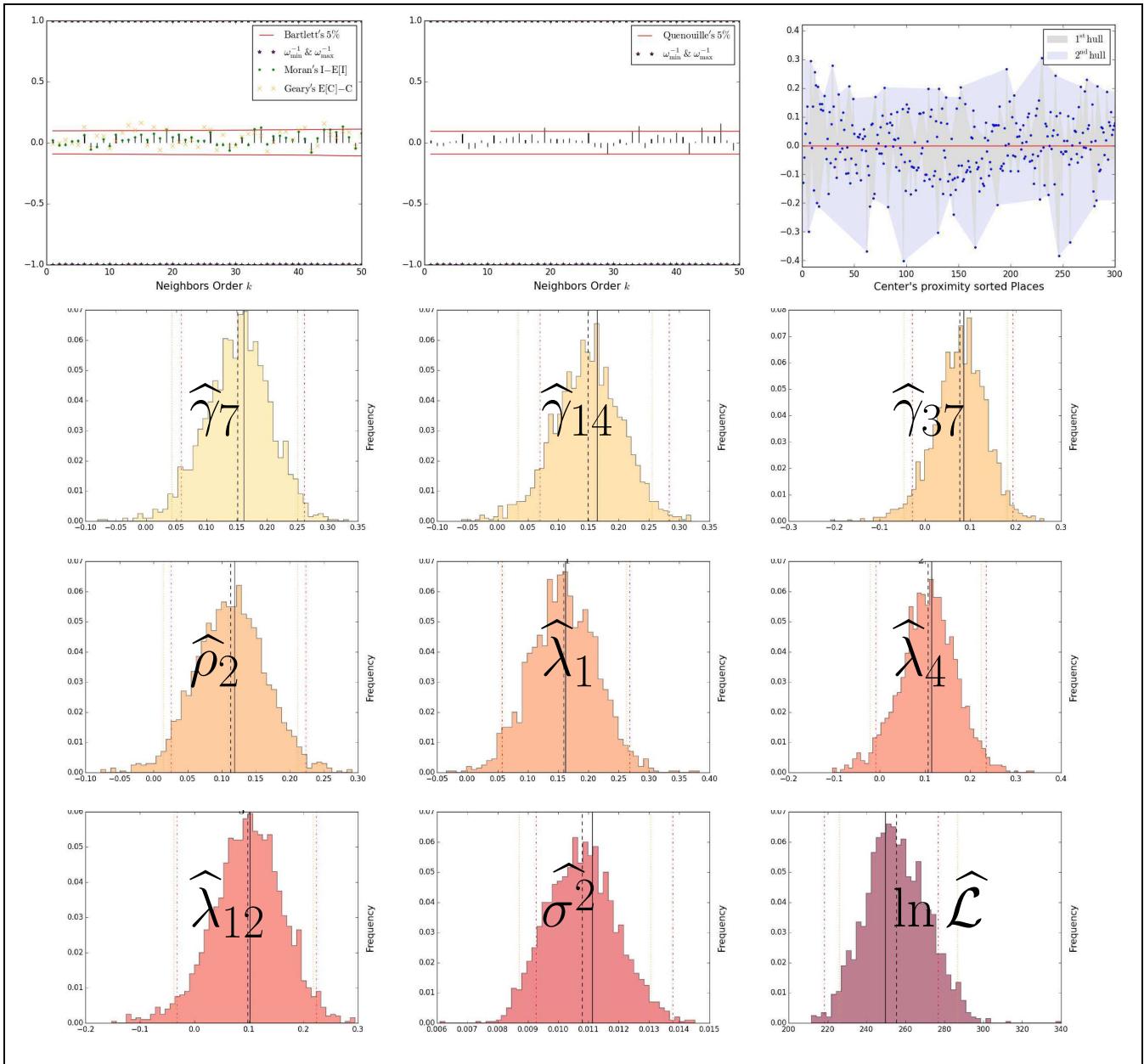


Figure 2.18: $\text{SARIMA}(\{2\}, \{7, 14, 37\}, \{1, 4, 12\})$'s (P)ACFs, residual plot and bootstrap-based density plots of spatial parameters (3rd rank)

Table 2.6: SARIMA ($\{2\}, \{7, 14, 37\}, \{1, 4, 12\}$)

Coeff.	Dist. ^a	Std. Err.	P> z	[95% C.I.]
β_0	D1	4.525859	0.000	-34.2063 -16.3909
	D2	.	.	.
	D3	4.651981	0.000	-35.3598 -16.7302
	D4	4.735276	0.000	-35.5907 -17.0287
β_1	D1	0.411140	0.000	1.6979 3.3163
	D2	.	.	.
	D3	0.424019	0.000	1.7352 3.4201
	D4	0.431791	0.000	1.7654 3.4580
β_2	D1	0.032517	0.000	0.2190 0.3470
	D2	.	.	.
	D3	0.032544	0.000	0.2226 0.3497
	D4	0.032271	0.000	0.2195 0.3460
γ_7	D1	0.054314	0.003	0.0543 0.2681
	D2	.	.	.
	D3	0.053774	0.011	0.0424 0.2506
	D4	0.051831	0.006	0.0580 0.2612
γ_{14}	D1	0.057506	0.004	0.0507 0.2770
	D2	.	.	.
	D3	0.055596	0.012	0.0339 0.2547
	D4	0.054434	0.003	0.0699 0.2833
γ_{37}	D1	0.058350	0.143	-0.0294 0.2003
	D2	.	.	.
	D3	0.057416	0.183	-0.0462 0.1811
	D4	0.056600	0.120	-0.0284 0.1935
ρ_2	D1	0.049549	0.016	0.0214 0.2164
	D2	.	.	.
	D3	0.051456	0.030	0.0142 0.2115
	D4	0.050322	0.016	0.0261 0.2234
λ_1	D1	0.049255	0.001	0.0652 0.2591
	D2	.	.	.
	D3	0.054161	0.003	0.0558 0.2631
	D4	0.053734	0.003	0.0575 0.2682
λ_4	D1	0.054839	0.036	0.0070 0.2229
	D2	.	.	.
	D3	0.062115	0.100	-0.0213 0.2236
	D4	0.061811	0.067	-0.0082 0.2341
λ_{12}	D1	0.052822	0.055	-0.0024 0.2055
	D2	.	.	.
	D3	0.064349	0.130	-0.0379 0.2175
	D4	0.065256	0.113	-0.0321 0.2237
δ^H	D1	.	.	.
	D2	.	.	.
	D3	0.057641	0.000	0.2717 0.4938
	D4	0.056590	0.000	0.2688 0.4907
δ^a	D1	.	.	.
	D2	0.011030	0.000	0.0794 0.1228
	D3	0.020867	0.000	0.0701 0.1522
	D4	0.020605	0.000	0.0684 0.1492
σ_r^2	D1	0.000920	.	.
	D2	.	.	.
	D3	0.001136	.	0.0087 0.0131
	D4	0.001149	.	0.0093 0.0138

Coeff.	Dist. ^a	P> z			
		$\mathcal{H}_0 : \hat{\cdot} = -1.00$	$\mathcal{H}_0 : \hat{\cdot} = 0.05$	$\mathcal{H}_0 : \hat{\cdot} = 0.12$	$\mathcal{H}_0 : \hat{\cdot} = 0.15$
γ_7	D3	0.0000	.	.	.
γ_{14}	D4	0.0000	.	.	.
γ_{37}	D3	0.0000	.	.	.
ρ_2	D4	0.0000	.	.	.
λ_1	D3	0.0000	.	.	.
λ_4	D4	0.0000	.	.	.
λ_{12}	D3	0.0000	.	.	.
δ^H	D3	.	0.0000	0.0104	0.8593
δ^a	D4	.	0.0000	0.0123	0.8930

^a C.I.-related objects derive from four types of distribution:

*D1 stands for normal-approximation-based distribution.

*D2 is as D1, but derives from an information matrix expressed in terms of δ^H and δ^a .

*D3 stands for bootstrap-based distribution. Associated std. are computed over bootstrap distributions.

*D4 stands for BCa bootstrap-based distribution. Associated std. are deduced by reversing the symmetry-based C.I. formula.

2.B Models development

This section has a computational scope and shows successively how to compute all the objects that are involved during the declination of the model from its simplest form into that of a SARIMA process. Let's start with the traditional linear model, as presented in eq.(2.23):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}$$

Under the assumption of unmodeled correlation structure for \mathbf{v} , *i.e.* uncorrelated and homoscedastic, its covariance matrix is $\mathbb{E}[\mathbf{v}\mathbf{v}'] = \sigma_v^2 \mathbf{I}$.

2.B.1 From the linear to the SEM

If one has good reasons to think that the above model should be turned into an autoregressive process in \mathbf{v} , *i.e.* that one should augment eq.(2.23) into eq.(2.26), one gets the following system to consider

$$\begin{cases} \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \\ \mathbf{v} = \mathbf{G}(\boldsymbol{\gamma})\mathbf{v} + \mathbf{u} \end{cases}$$

which leads to

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{G}(\boldsymbol{\gamma})\mathbf{v} + \mathbf{u} \\ &= \mathbf{X}\boldsymbol{\beta} + \mathbf{G}(\boldsymbol{\gamma})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \mathbf{u} \\ \Leftrightarrow \mathbf{u} &= (\mathbf{I} - \mathbf{G}(\boldsymbol{\gamma}))(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \boldsymbol{\Gamma}(\boldsymbol{\gamma})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ \Leftrightarrow \mathbf{y}_\gamma &= \mathbf{X}_\gamma\boldsymbol{\beta} + \mathbf{u} \end{aligned}$$

Where the above line is uniquely defined only if $\boldsymbol{\Gamma}(\boldsymbol{\gamma})$ is invertible. If so, and assuming that \mathbf{u} has no explicitly modelled covariance, it follows that that of \mathbf{v} now becomes

$$\begin{aligned} \mathbb{E}[\mathbf{v}\mathbf{v}'] &= \mathbb{E}[\boldsymbol{\Gamma}(\boldsymbol{\gamma})^{-1}\mathbf{u}\mathbf{u}'\boldsymbol{\Gamma}(\boldsymbol{\gamma})^{-1'}] \\ &= \sigma_u^2 \left(\boldsymbol{\Gamma}(\boldsymbol{\gamma})'\boldsymbol{\Gamma}(\boldsymbol{\gamma}) \right)^{-1} \\ &= \boldsymbol{\Omega}_\gamma \end{aligned}$$

2.B.2 From the SEM to SARIMA

If one then has good reasons to think that the above model should in turn be turned into an autoregressive process in \mathbf{y}_γ with a moving average process in \mathbf{u} , *i.e.* that one should augment

eq.(2.26) into eq.(2.28), the system to consider is

$$\begin{cases} \mathbf{y}_\gamma = \mathbf{Q}(\boldsymbol{\rho})\mathbf{y}_\gamma + \mathbf{X}_\gamma\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} = \mathbf{L}(\boldsymbol{\lambda})\mathbf{r} + \mathbf{r} \end{cases}$$

which leads to

$$\begin{aligned} \mathbf{y}_\gamma &= \mathbf{Q}(\boldsymbol{\rho})\mathbf{y}_\gamma + \mathbf{X}_\gamma\boldsymbol{\beta} + \mathbf{L}(\boldsymbol{\lambda})\mathbf{r} + \mathbf{r} \\ \Leftrightarrow \mathbf{y}_{\gamma,\rho} &= \mathbf{X}_\gamma\boldsymbol{\beta} + \Lambda(\boldsymbol{\lambda})\mathbf{r} \end{aligned}$$

If the above equality is uniquely defined, it follows that

$$\begin{aligned} \mathbf{r} &= \Lambda(\boldsymbol{\lambda})^{-1} (\mathbf{y}_{\gamma,\rho} - \mathbf{X}_\gamma\boldsymbol{\beta}) \\ \Leftrightarrow \mathbf{y}_{\gamma,\rho,\lambda} &= \mathbf{X}_{\gamma,\rho}\boldsymbol{\beta} + \mathbf{r} \end{aligned}$$

If one assumes that \mathbf{r} has been totally filtered, *i.e.* $\mathbb{E}[\mathbf{r}\mathbf{r}'] = \sigma_r^2 \mathbf{I}$, it follows that the covariance matrices respectively of \mathbf{u} and \mathbf{v} are

$$\begin{aligned} \mathbb{E}[\mathbf{u}\mathbf{u}'] &= \mathbb{E}[(\mathbf{Q}(\boldsymbol{\rho})\mathbf{y}_\gamma + \Lambda(\boldsymbol{\lambda})\mathbf{r})(\mathbf{Q}(\boldsymbol{\rho})\mathbf{y}_\gamma + \Lambda(\boldsymbol{\lambda})\mathbf{r})'] \\ &= \mathbf{Q}\text{cov}[\mathbf{y}_\gamma, \mathbf{y}_\gamma]\mathbf{Q}' + 2\mathbf{Q}\mathbb{E}[\mathbf{y}_\gamma\mathbf{r}']\Lambda' + \sigma_r^2\Lambda\Lambda' \\ &= \Omega_{\gamma,\rho,\lambda} \\ \mathbb{E}[\mathbf{v}\mathbf{v}'] &= \Gamma^{-1}\Omega_{\gamma,\rho,\lambda}\Gamma^{-1'} \end{aligned}$$

References

- Allio, C. (2015). "Local Policies, Urban Dynamics and Climate Change: Development of a Multiscale Modeling Framework". PhD thesis, AllioT2015.
- Anselin, L (1982). "A Note on Small Sample Properties of Estimators in a First-Order Spatial Autoregressive Model". *Environment and Planning A* 14, pp. 1023–1030.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, p. 289.
- Anselin, L. (2002a). "Under the hood:Issues in the specification and interpretation of spatial regression models". *Agricultural Economics* 27, pp. 247–267.
- Anselin, L. (2002b). "Under the hood:Issues in the specification and interpretation of spatial regression models". *Agricultural Economics* 27, pp. 247–267.
- Anselin, L. (2007). "Spatial econometrics in RSUE: Retrospect and prospect". *Regional Science and Urban Economics* 37.4, pp. 450–456.
- Anselin, L. (2010). "Thirty years of spatial econometrics". *Papers in Regional Science* 89.1, pp. 3–25.
- Anselin, L. (2011). *GMM Estimation of Spatial Error Autocorrelation with and without Heteroskedasticity*. Tech. rep. GeoDa Center for Geospatial Analysis and Computation – Arizona State University, pp. 1–23.
- Anselin, L. and S. J. Rey (2007). "PySAL: A Python Library of Spatial Analytical Methods". *The Review of Regional Studies* 37.1, pp. 5–27.
- Anselin, L. and others (1980). "Estimation methods for spatial autoregressive structures." *Regional Science Dissertation & Monograph Series, Program in Urban and Regional Studies, Cornell University* 8, p. 273.
- Armington, P. S. (1969). "A Theory of Demand for Products Distinguished by Place of Production". *Staff Papers (International Monetary Fund)* 16.1, pp. 159–178.
- Arraiz, I., D. M. Drukker, H. H. Kelejian and I. R. Prucha (2010). "A spatial cliff-ord-type model with heteroskedastic innovations: Small and large sample results". *Journal of Regional Science* 50.2, pp. 592–614.
- Askenazy, P. and X. Timbeau (2003). "Partage de la valeur ajoutée et rentabilité du capital en France et aux États-Unis : une réévaluation". *Economie et statistique* 363.1, pp. 167–189.
- Autant-Bernard, C. and J. P. LeSage (2011). "Quantifying Knowledge Spillovers Using Spatial Econometric Models". *Journal of Regional Science* 51.3, pp. 471–496.
- Badinger, H. and P. Egger (2011). "Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances". *Papers in Regional Science*.
- Behrens, K. and F. Robert-Nicoud (2011). "Tempora mutantur: In search of a new testament for NEG". *Journal of Economic Geography* 11.2, pp. 215–230.
- Blanchflower, D. G. and A. J. Oswald (1995). "The Wage Curve". *Journal of Economic Literature* 33.2, pp. 785–799.

- Bodson, P. and D. Peeters (1975). "Estimation of the Coefficients of a Linear Regression in the Presence of Spatial Autocorrelation. An Application to a Belgian Labour-Demand Function". *Environment and Planning A* 7.4, pp. 455–472.
- Bosi, S., E. Iliopoulos and H. Jayet (2010). "Optimal Immigration Policy When the Public Good Is Rival".
- Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*.
- Bradley, E. (1979). "Bootstrap Methods: Another Look at the Jackknife". *The Annals of Statistics* 7.1, pp. 1–26.
- Breusch, T. S. and A. R. Pagan (1979). "A Simple Test for Heteroscedasticity and Random Coefficient Variation". *Econometrica*.
- Brigitte Baccaïni, C. D. (2007). "Inter-Regional Migration Flows in France over the Last Fifty Years". *Population (English Edition, 2002-)* 62.1, pp. 139–155.
- Calatayud, P., J. Duvey-Pilate and D. Malody (2011). *Bilan économique et social de la région Centre*. Tech. rep. INSEE, p. 45.
- Choi, E. J. and J. Spletzer (2012). "The declining average size of establishments: evidence and explanations". *Monthly Labor Review* March, pp. 50–65.
- Cliff, A. D. and J. K. Ord (1973). *Spatial Autocorrelation*. Pion, London.
- Corrado, L. and B. Fingleton (2012). "Where is the economics in spatial econometrics?" *Journal of Regional Science* 52.2, pp. 210–239.
- Cressie, N. A. C. (1993). "Statistics for Spatial Data". *Computational Statistics & Data Analysis* 14.4, p. 547.
- Das, D., H. H. Kelejian and I. R. Prucha (2003). "Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances". *Papers in Regional Science* 82.1, pp. 1–26.
- Davidson, R. and E. Flachaire (2008). "The wild bootstrap, tamed at last". *Journal of Econometrics* 146.1, pp. 162–169.
- Dechezleprêtre, A., R. Martin and M. Mohnen (2013). "Knowledge spillovers from clean and dirty technologies: A patent citation analysis". *Grantham Research Institute and the Environment Working Paper No 151* December, pp. 1–47.
- Dogan, O. (2013). "Heteroskedasticity of Unknown Form in Spatial Autoregressive Models with Moving Average Disturbance Term". *Journal of Econometrics* December, pp. 101–127.
- Drukker, D. M., P. Egger and I. R. Prucha (2013). "On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors". *Econometric Reviews* 32, pp. 686–733.
- Durand, D. (2012). *Le marché du logement à Paris 2001-2011: retour sur dix ans d'accession*. Tech. rep., p. 22.
- Efron, B and R. J. Tibshirani (1993). "An introduction to the bootstrap". *Refrigeration And Air Conditioning*.

- Elhorst, J. P., D. J. Lacombe and G. Piras (2012). "On model specification and parameter space definitions in higher order spatial econometric models". *Regional Science and Urban Economics* 42.1-2, pp. 211–220.
- Elhorst, J. (2001). "Dynamic models in space and time". *Geographical Analysis* 33.2, pp. 119–140.
- Elhorst, P and S. Vega (2013). "On spatial econometric models, spillover effects, and W". *ERSA conference papers* February, pp. 1–28.
- Fingleton, B. (2009). "Spatial autoregression". *Geographical Analysis* 41.4, pp. 385–391.
- Gaspar, M. J. L. (2017). "New Economic Geography: Perspectives, multiple regions and individual heterogeneity". PhD thesis, p. 182.
- Geary, R. C. (1954). "The Contiguity Ratio and Statistical Mapping". *The Incorporated Statistician*.
- Gibbons, S. and H. G. Overman (2012). "Mostly pointless spatial econometrics?" *Journal of Regional Science* 52.2, pp. 172–191.
- Gonçalves, S. and D. Politis (2011). "Discussion: Bootstrap methods for dependent data: A review". *Journal of the Korean Statistical Society* 40.4, pp. 383–386.
- Guidetti, F. and INSEE (2012). *En 2011, la consommation des ménages marque le pas*. Tech. rep., pp. 1–4.
- Haggett, P. (1965). *Locational analysis in human geography*. Edward Arnold, p. 339.
- Hannan, E. J. and B. G. Quinn (1979). "The Determination of the Order of an Autoregression". *Journal of the Royal Statistical Society*.
- Hungerland, W.-F. (2017). "The Gains from Import Variety in Two Globalisations: Evidence from Germany". *Institute of Economic History, School of Business and Economics, Humboldt-University of Berlin*, mimeo.
- Joly, I. and Y. Crozet (2006). "Budgets temps de transport et vitesses : de nouveaux enjeux pour les politiques de mobilité urbaine". *La ville aux limites de la mobilité Actes du colloque "La ville aux limites de la mobilité"*; Paris; 2006. Ed. by M. BONNET and P. AUBERTEL. Presses Universitaires de France, pp. 287–296.
- Joly, I., Y. Crozet, P. Bonnel and C. Raux (2002). *La "Loi" de Zahavi, quelle pertinence pour comprendre la contraction ou la dilatation des espaces-temps de la ville ?* Tech. rep.
- Kaiser, U. (2002). "Measuring knowledge spillovers in manufacturing and services: an empirical assessment of alternative approaches". *Research Policy* 31.1, pp. 125–144.
- Kelejian, H. H. and I. R. Prucha (1999). "A generalized moments estimator for the autoregressive parameter in a spatial model". *International Economic Review*.
- Kelejian, H. H. and I. R. Prucha (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances". *Journal of Econometrics* 157.1, pp. 53–67.
- Kunsch, H. R. (1989). "The Jackknife and the Bootstrap for General Stationary Observations". *The Annals of Statistics* 4.2, pp. 1217–1241.

- Lee, L.-f. (2003). "Best Spatial Two Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances". *Econometric Reviews* 22.4, pp. 307–335.
- Lee, L. F. (2004). "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Econometric Models". *Econometrica* 72, pp. 1899–1926.
- Lee, L.-f. and X. Liu (2010). *Efficient Gmm Estimation of High Order Spatial Autoregressive Models With Autoregressive Disturbances*. Vol. 26. 01, pp. 187–230.
- LeSage, J. and R. K. Pace (2009). *Introduction to Spatial Econometrics*, p. 374.
- LeSage, J. P. (1999). "The Theory and Practice of Spatial Econometrics". *International Journal of Forecasting* 2.2, pp. 245–246.
- Lin, K.-p., Z.-h. Long and W. Mei (2007). "Bootstrap Test Statistics for Spatial Econometric Models".
- Martellosio, F. (2011). "Efficiency of the OLS estimator in the vicinity of a spatial unit root". *Statistics and Probability Letters* 81.8, pp. 1285–1291.
- McDaniel, C. a. and E. Balistreri (2003). "A review of Armington trade substitution elasticities". *Economie internationale* 2.94/95, pp. 301–313.
- Moran, P. A. P. (1950). "Notes on Continuous Stochastic Phenomena". *Biometrika*.
- Morer, N. and INSEE (2015). *La consommation des ménages est encore convalescente en 2014*. Tech. rep., pp. 1–4.
- Nelder, J. and R. Mead (1965). "A simplex method for function minimization". *The computer journal*.
- Ottaviano, G. I. P. and P. Martin (2001). "Growth and Agglomeration". *International Economic Review* 42.4, pp. 947–968.
- Paelinck, J. and P. Klaassen (1979). *Spatial econometrics*. Saxon House, Farnborough.
- Pinkse, J. and M. E. Slade (2010). "The future of spatial econometrics". *Journal of Regional Science* 50.1, pp. 103–117.
- Rivera-Batiz, F. (1988). "Increasing returns, monopolistic competition, and agglomeration economies in consumption and production". *Regional Science and Urban Economics* 18.1, pp. 125–153.
- Samuelson, P. A. (1952). "The Transfer Problem and Transport Costs: The Terms of Trade When Impediments are Absent". *The Economic Journal* 62.246, pp. 278–304.
- Shapiro, S. S. and M. B. Wilk (1965). "An Analysis of Variance Test for Normality (Complete Samples)". *Biometrika* 52.3/4, pp. 591–611.
- U.S. Bureau of Labor Statistics (2012). *Consumer Expenditure in 2010: Lingering Effects of the Great Recession*. Tech. rep. August. US Bureau of Labor Statistics.
- U.S. Bureau of Labor Statistics (2013). *Consumer Expenditures in 2011*. Tech. rep. April, pp. 1–23.
- Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. London : Pion, p. 166.
- Zahavi, Y. (1979). *The 'UMOT' Project*. Tech. rep. l'U.S. Department of Transportation and the Ministry of Transport of Federal Republic Of Germany.

Zahavi, Y. and A. Talvitie (1980). "Regularities in travel time and money expenditures". *59th Annual Meeting of the Transportation Research Board*. 750.