Proximity or opportunity? Spatial and market determinants of private individuals' buy-to-let investments

Antoine Peris, Laure Casanova Enault

Proximity or opportunity? Spatial and market determinants of private individuals' buy-to-let investments

Antoine Peris*†
antoine.peris@univ-avignon.fr

Laure Casanova Enault*
laure.casanova@univ-avignon.fr

November 23, 2023

Abstract

This paper contributes to the debate on the liquidity of real estate investment in the context of financialisation. Using microdata built from tax registers, we analyse the geography of rental housing purchases by private individuals from three French cities. We develop a modelling approach in order to better understand the respective roles of space and market characteristics in determining buy-to-let investment flows. Considering the distribution of the data and our objective of integrating both intra- and intercity housing investments in a single model, we use an adaptive zoning approach. This approach allows high spatial resolution where interactions are strong to be kept, and the aggregation of more distant, less populated areas. We demonstrate that geographical proximity is highly determinant in explaining flows of buy-to-let investments from private individuals. We also uncover striking facts related to the geography of rental investments, such as the convergence of investments from rich suburbs toward the centre of agglomerations and preferential flows from the Paris region to southern and coastal cities. Finally, we find that investors tend to buy in upmarket areas and in places that are more expensive than their market of residence. Our results indicate that geographical proximity and safety of investments are key factors in housing wealth accumulation by private individuals.

Key words: Multiple property ownership, rental investment, register data, spatial interaction models, adaptive zoning.

*UMR CNRS 7300 ESPACE, Avignon Université, F-84029 Avignon, France
†Corresponding author

We would like to thank members of the WIsDHoM team, especially Guilhem Boulay, Renaud Le Goix and Loïc Bonneval for the discussions and comments throughout this research. We are also grateful to Rémi Lei and Martin Bocquet for their contribution to the data enhancement work and for sharing their expertise on land registry and transaction databases. We also thank the DGFiP, DGALN and CEREMA for providing the data. This work was supported by the Agence Nationale de la Recherche (France) under grant ANR-18-CE41-0004 (WIsDHoM project, Wealth inequalities and the dynamics of housing markets).

1 Introduction

Today, housing is a major driver of inequalities. Decades of house price rises in OECD countries have led to a growing divide between people that own residential units and people that are priced out of the market (Adkins et al., 2021). While younger and lower-income households have progressively been excluded from home-ownership (Arnold and Boussard, 2017; Lennartz et al., 2016) especially in the central areas of large cities (Le Goix et al., 2021), other social groups have benefited from the resulting rental boom by becoming multiple property owners (Kadi et al., 2020; Ronald and Kadi, 2018).

Over the last few years, information on the concentration of housing wealth has started to become available in several countries. Studies have documented the prominent position of private individuals owning multiple properties in countries such as the United-Kingdom (Ronald and Kadi, 2018), the Netherlands (Hochstenbach, 2022) or France (André and Meslin, 2021). While these studies give very valuable information on the magnitude of the phenomenon and the socio-demographic characteristics of landlords, the spatial dimension of multiple property ownership remains understudied.

It is only very recently that a burgeoning literature has started to look at the geography of investment in housing by private individuals (Hochstenbach, 2023; Levy, 2021). Understanding the spatial strategy of private rental investors is all the more important because landlords have a strong impact on local housing markets: they are spatially sorting tenants (Paccoud, 2017; Rosen, 2014), they may participate in eviction mechanisms (Shelton, 2018), and increase competition in high-demand areas, potentially leading to higher house prices (Chinco and Mayer, 2016; Hulse and Reynolds, 2018).

The literature on investment in housing has argued for an increasing liquidity of real estate investment in a context of financialisation (Beswick et al., 2016; Fernandez and Aalbers, 2016; Fields, 2022; Taşan-Kok et al., 2021). This context would lead to investment practices that are guided mainly by the characteristics of the market, regardless of its geographical location, which would favour the development of long-distance investments. While this might be true for large corporate players, recent research studies have shown that households and individuals investing in housing tend to be 'home-biased', meaning that they invest locally or regionally (Hochstenbach, 2023; Levy, 2021). An in-depth analysis of the behaviour of this type of landlord is essential, given their importance in France (Casanova Enault et al., 2023) and given the revival of private landlordism over the last two decades (Ronald and Kadi, 2018; Wind et al., 2020).

This paper contributes to the debate on the liquidity of real estate investment by looking specifically at the spatial dimension of individuals' and households' investments in rental housing. In order to go beyond mapping and descriptive statistics, we have developed a modelling approach enabling a better understanding of the respective roles of space and market characteristics in determining investment flows in rental housing. We use spatial interaction models to analyse the role of different factors as well as to highlight preferential flows of investment within and between cities. We pay special attention to territorial heterogeneity by focusing on households living in – and investing from – three cities with very different local market characteristics and positions in the French urban hierarchy: the Paris region (the capital city), Lyon (a regional centre), and Avignon (a medium-sized city). In order to perform this analysis, we created a dataset by integrating and enriching microdata from tax registers and real estate transaction deeds covering a population of 8.2 million individual owners. Focusing on owner-occupiers from the three cities, we tracked 223,095 investments in rental housing made by 188,963 different households between 2010 and 2018.

Using these data in a spatial interaction framework led to a methodological challenge. The origin-destination matrix of rental investment is sparse: few very large flows between some areas and many very small flows. Three solutions are commonly applied in this situation : i) aggregating flows into wider zones – thus losing information on investment in proximity; ii) removing small flows – thus losing complexity in the data and biasing the estimation; iii) separating short and long distance investment by using two models – thus assuming that they would obey different logics. To avoid these pitfalls and to make the most of the wealth of our detailed data, we adopted an adaptive zoning approach (Hagen-Zanker and Jin, 2012). The idea of this approach is to keep high spatial resolution where interaction is strong (for example at short distances or between areas of a significant size) and to aggregate faraway and less populated places where spatial interactions happen less frequently. Such an adaptive zoning strategy is appropriate to represent investors' geographical perception and allows the investigation of multiple spatial scales in housing investment, from intra-urban investments to inter-urban ones, into a single model.

The following section (2) reviews previous research into multiple property ownership and investment distance. Then we present the database on investment flows that we built from microlevel land registry and transaction data, as well as some descriptive statistics on these investments (3). The fourth section presents our methodological choices for modelling investment flows within and between local markets (4). Next, we present the results of our modelling approach (5). The last section gives the conclusions of this study.

2 Rental boom and the geography of real estate investment

2.1 Post-homeownership society and the resulting rental boom

Over the last few years, many OECD countries have experienced rising wealth inequalities, especially regarding housing. In France the orientation toward asset-based welfare has been characterized by two main housing policies targeting households from the middle and upper classes: an easier access to homeownership and support for buy-to-let investments (Benites-Gambirazio and Bonneval, 2022). Promotion of homeownership has been a central feature of housing policies in France for decades in a context of real estate price inflation driven by low interest rates and long loan periods (Le Goix et al., 2021). Between 1954 and 2018, the homeownership rate increased from 35% to 58% (Bonvalet and Bringé, 2013; INSEE, 2020). In parallel, French housing policies fostered the private rental sector through tax incentives and subsidies (Vergriete and Guerrini, 2012). Housing wealth increased by 133% in France between 1998 and 2015 (Ferrante and Solotareff, 2018), and is mostly concentrated in the hands of private individuals (Casanova Enault et al., 2023).

However, since the Global Financial Crisis (GFC), the homeownership rate in France has stabilised and is even decreasing for certain categories of the population such as lower-income and younger households (Arnold and Boussard, 2017). At the same time, the private rental market in the country has expanded in the country (INSEE, 2020). This shift is occurring in several OECD countries and has been conceptualised as a transition towards a "post-homeownership society" (Ronald, 2008). This transition has been accompanied by a "rental boom" and a rise in multiple property ownership, mostly among middle-aged and affluent households (Kadi et al., 2020). According to Aalbers et al. (2021), this situation comes from the fact that financialisation through mortgage debt has reached its limits. These trends of a declining homeownership rate and rising inequalities in the concentration of housing wealth in countries that actively promoted access to property have been described by Arundel and Ronald (2021) as the "false promise of homeownership". In France, a recent study has documented the concentration of housing wealth in the hands of the most affluent households. According to André and Meslin (2021), the number of homes owned increases with income. While 67% of households in the first two income deciles do not own a house, 60% of households from the last decile are multiple property owners. Regarding large multiple property owners, the concentration is even stronger: among the top 1%, 33% own more than 5 housing units, 42% for the top 0.1%. For less affluent households, access to homeownership and, even more, to multiple property ownership is more and more difficult. This was shown by Piketty (2019) with long-term data on the divergent dynamic of housing wealth concentration.

2.2 Spatial anchoring of real estate investments

Knowing the spatial dimension of private individuals' investment strategy is essential given their importance in housing markets. By selecting certain types of areas, landlords will influence the spatial sorting of tenants (Rosen, 2014), increase competition with other market actors or influence price with speculative behaviours (Chinco and Mayer, 2016). However, the spatial dimension of landlords activities has been insufficiently scrutinised.

Urban studies literature has focused on understanding how financialisation "traverses geographical [...] scales" (Fields, 2017). As shown by Özogul and Tasan-Kok (2020) in their meta-analysis, a significant number of the papers studied differentiate residential property investors depending on their scale of operation. In most cases, studies distinguish between local and international investors, but there are differences in the way these scales are defined. Transnational flows of real estate investment at the top of the urban hierarchy have received a lot of attention over the past years. Studies have documented the rising importance of large transnational corporate landlords in global cities (Fields and Uffer, 2016; Taşan-Kok et al., 2021) or the cross-border investments by rich and super-rich individuals (Alstadsæter et al., 2022; Paris, 2013). It has also been shown that the rise of digital technologies allows landlords to build and manage large dispersed portfolios (Fields, 2022). These works tend to show that, for certain market actors, property investment is becoming an increasingly deterritorialised practice.

It is unclear how this tendency is affecting private individuals, who remain the largest group of owners in many countries such as France, the United Kingdom, or the Netherlands. One could argue that in a context of digitalisation and commodification of housing, investors may buy housing units in areas where rental profit or capital gains are maximized, no matter what the distance. In the French context, this idea has been supported by Vergriete and Guerrini (2012) who showed that in areas targeted by place-based government policies enabling tax reductions when investing in rental real estate, new types of investors that are 'emotionally and geographically' distant from the homes they purchase have emerged. These findings are supported by an increase in the average distance between the place of residence and the place of investment of private individuals targeting such areas (Vergriete, 2013).

However, in contradiction with the tendency observed for this group of investors and for large corporate players, studies focusing on private individuals have argued that households investing in buy-to-let properties are 'home-biased', meaning that they invest locally or regionally (Hochstenbach, 2023; Levy, 2021). This can be explained by the fact that they give priority to targeting markets they already know, to be sure they are paying the right price[1] and to reduce costs related to the management of their portfolio.

[1]This idea of decrease of the quantity and quality of information with distance has been explored in

3 Building a database on investments in rental housing

3.1 Available data

We created a dataset on rental investment by integrating two databases: a first one describing the property rights of individuals as well as the characteristics of properties (*Fichiers Fonciers*), and a second one covering real estate transactions (*DV3F*)[2]. More information on the databases and on the steps required to build the dataset is provided in the Data section of the Supplementary Material. The result is a comprehensive dataset on investments in rental housing by households living in the FUAs of Paris, Lyon, and Avignon[3]. The dataset covers the period 2010-2018 and is geocoded at both the place of residence and at the place of investment, and enables a study of capital mobility from an origin-destination perspective.

3.2 Summary statistics on investments

In total, we tracked 223,095 buy-to-let investments by 188,963 different owner-occupier individuals and households. Table 1 presents descriptive statistics about these investments. A large share originates from the FUA of Paris and represents 84% of the total number. Lyon represents 14% and Avignon 2%.

With a median investment distance of 10.3 km, it appears that most investors buy properties that are located close to their place of residence. However, there are pronounced differences between the three cities. The individuals and households from Lyon are the ones that invested the closest, with 69.5% of purchases that took place within the FUA and 50% at less than 8.9 km. Investors from Avignon and Paris seem more outward-looking in comparison, with a share of investments contained within the FUA of respectively 59.3% and 63.6%. These two cities differ in terms of investment distance. Parisians tend to buy rental units located further away, with 50% of the investments targeting dwellings at less than 11.6 km (9.5 km for Avignon) and 25% of investments at more than 116.6 km (42.1 km for Avignon).

economics literature through the concept of information asymmetry (Qiu et al., 2020). It is also consistent with Hägerstrand's explanation of the role of distance in reducing spatial interaction (Hagerstrand et al., 1968).

[2] Both datasets are provided by DGFiP-DGALN-CEREMA.

[3] These investments can be located anywhere in mainland France, except in the départements of Haut-Rhin, Bas-Rhin and Moselle. Due to the application of local law, information on transactions in these areas is not directly transmitted to the Public Finances Directorate General (DGFiP) that centralises the data.

Table 1: Descriptive statistics of buy-to-let investments from the studied FUAs

	Type	Freq.	Share (%)	Road distance (km) median	Road distance (km) Q3	Road distance (km) mean	Duration (minutes) median	Duration (minutes) Q3	Duration (minutes) mean
Paris	intra-FUA	119 338	63.6	11.6	116.6	89.1	16.5	101.4	67.2
	extra-FUA	68 190	36.4						
Lyon	intra-FUA	21 547	69.5	8.9	42.1	58.9	13.6	43.7	47.6
	extra-FUA	9 468	30.5						
Avignon	intra-FUA	2 700	59.3	9.5	54.0	61.2	14.2	57.5	51.3
	extra-FUA	1 852	40.7						
Total	-	223,095	-	10.3	88.6	79.8	15.3	77.5	61.2

Note: Road distances and trip durations were computed with the R osrm package (Giraud, 2022) that uses OpenStreetMap data.

4 An adaptive market spatial interaction model

4.1 Modelling investment flows with spatial interaction models

Using a spatial interaction model appears to be the most interesting framework to model the impact of space and market characteristics on investment flows. Such a meso-scale model, also known as gravity model, explains the magnitude of flows between two areas based on the distance between them and the characteristics of the emitting and receiving zones. This kind of model is widely used in geography, economics, and network science to study human mobility (Pumain, 1986), information circulation (Peris et al., 2021), and international trade (De Benedictis and Taglioni, 2011). Recently it has also been used to model real estate investment flows (Alstadsæter et al., 2022; Levy, 2021; McAllister and Nanda, 2016; Zhang et al., 2020). However, these previous applications to real estate investments used large spatial aggregates such as countries, regions or large commuting zones.

In our case, using a finer spatial resolution led to a methodological challenge. Once aggregated into an origin-destination matrix, the flows of investments between locations exhibit a strongly asymmetric distribution. There are few large flows and many small ones, most of them not reaching the confidentiality threshold requested by the French tax service for the presentation of the results ($F_{ij} \geq 11$). Moreover, the initial matrix was filled with many zeros, which makes it difficult to produce a robust estimation of the parameters of the gravity equation (Burger et al., 2009). Such a distribution comes from the fact that investments are relatively rare events both in time and space. When observed at a fine spatial scale, they do not meet the requirement of normality necessary to apply the model. Different solutions exist to make an origin-destination matrix denser. Flows could be measured over a longer period of time – something we already do here by working on the period 2010–2018. Alternatively, wider zones could be used for the aggregation.

However, that would cause the loss of a great deal of information on investments, since we have seen previously that a lot of them take place in the vicinity of the investor's place of residence. Calibrating the model on data points reaching minimum thresholds would also be problematic, because removing small and zero flows could bias the estimation. Finally, separating the estimation of the gravity equation into two models with one for intraurban flows and one for interurban flows would also have been an unsatisfactory solution, because our objective was to understand the spatial and market determinants of investment flows no matter what the scale of operation.

4.2 An adaptive zoning strategy

Adaptive zoning appeared to be the solution to integrate the multiple spatial scales of investment — from intra- to interurban — into a single model and to make the most of the very detailed geographic information in our dataset. This approach allowed us to keep a fine spatial resolution where interaction is strong (i.e. at short distances, or between large-sized areas) and to merge faraway and less populated places where spatial interactions happen less frequently. While this method was initially designed for optimizing computations (Hagen-Zanker and Jin, 2012), it also presents an interest from a conceptual point of view. As stated by Cottineau et al. (2019), "such adaptive zoning is probably more appropriate to represent human geographical perception than the uniform mesh sizes generally used in territorial models". In the case of real estate investment, it is likely that a household willing to invest in buy-to-let housing will view a nearby neighbourhood as a coherent spatial entity while a city 600 km away will be imagined as a more fuzzy area, including for example the whole metropolitan area. Such zoning systems allowing the observation of both intra- and interurban flows also make sense in a context where these two scales of analysis are increasingly blurred (Cardoso and Meijers, 2021).

The adaptive zoning method relies on two main steps. The first one is a stepwise hierarchical aggregation procedure of all atomic zones into clustered zones. The second is a process of neighbourhood creation focusing on each atomic zone. At this step, some clustered zones will be disintegrated, based on mass and distance criteria, in order to create a distance varying zoning system for each emitting zone. This section presents a simple implementation of the method initially developed by Hagen-Zanker and Jin (2012) for spatial interaction modelling based on commuting data.

Our algorithm starts from a vector of zone populations, a vector of zone areas, a distance matrix between the centroids of our atomic zones, and a contiguity matrix. For the second phase of the algorithm that will be described below, we also include a vector of internal distance for each zone. Internal distance is the average distance between two points in an area. It is defined as $\sqrt{a_i/\pi}$ where a_i is the area of zone i.

In the first phase, i.e. the stepwise hierarchical aggregation, contiguous pairs of zones are merged iteratively when they minimize the merging criterion c_{ij} that combines both size and distance:

$$c_{ij} = d_{ij}(m_i + m_j)$$

where d_{ij} is the distance between i and j, m_i is the population of i and m_j the population of j.

At every step, the zone population and areas are adjusted to the new situation and a new distance matrix is computed:

$$d_{Ij} = \frac{\sum_{i \in I} m_i d_{ij}}{\sum_{i \in I} m_i}$$

where d_{Ij} corresponds to the distance between zone I – resulting from the merger — and zone j. Internal distances are adjusted in a similar way to take into account the location of the most populated elementary units:

$$d_{II} = \frac{\sum_{i \in I} \sum_{j \in I} m_i m_j (2d_{ij})}{2 \sum_{i \in I} \sum_{j \in I} m_i m_j}$$

This process is carried out until a single cluster groups all the original units. Panel A of Figure 1 presents three different iterations of this process.

The hierarchy of zone aggregations is then used for the second phase of the algorithm that creates distance varying zoning systems for different focus zones. Still following the Hagen-Zanker and Jin (2012) method, in this phase the algorithm incrementally breaks clustered zones based on a function δf_{ij} estimating the error in spatial interaction between two zones caused by the aggregation. This function takes the following form:

$$\delta f_{ij} = (m_i m_j e^{-\beta d_{ij}})(1 - e^{-\beta \delta d_{ij}})$$

where β is a parameter to be calibrated and δd_{ij} is the error in distance due to aggregation, and corresponds to the average intrazonal distance:

$$\delta d_{ij} = \frac{d_{ii} + d_{jj}}{2}$$

where d_{ii} is the intrazonal distance of i and d_{jj} is the intrazonal distance of j. This algorithm runs until it reaches the required number of zones.

The original units we used for the adaptive zoning algorithm are the *Établissements publics de coopération intercommunale* (EPCI), a territorial division that groups several municipalities with a common territorial development project[4]. The choice of the EPCIs was motivated by a trade-off between keeping a high spatial resolution and limiting the number of original units in order to reduce computation time for the adaptive zoning. Although we used a purely institutional division, it is at this scale that housing policies are implemented. Moreover, recent research into the Modifiable Areal Unit Problem (MAUP) related to the prediction of house prices has shown that this scale is among the most coherent of those used to study local housing markets in France (Josselin et al., 2023). The initial zoning contains 1220 spatial units, covering the entire territory of mainland France. As the EPCIs do not perfectly match the geography of the FUAs for which we have data, we kept only the spatial units fully contained within the FUAs as origins of the investment flows. This selection process resulted in 66 spatial units for the Paris urban area, 11 for Lyon, and 2 for Avignon.

The last step was to run tests for selecting both β and the number of clusters. We had two predefined criteria for the selection of the zoning systems: 1) big cities — even far away — should not be merged with rural surroundings, but remain coherent urban entities; 2) the share of flows in origin-destination pairs where $F_{ij} \geq 11$ should be as close as possible to 95% in order to present residuals on data points that respect the confidentiality threshold requested by our data provider. We generated zoning systems for the centre of the Paris FUA with different values of β according to the sequence $\beta_n = 0.002n - 0.001$ with $0.001 \geq n \leq 0.04$. Figure S1 in the Supplementary Material presents the resulting zoning systems for different values of β and a number of cluster $C = 225$. After visual inspection, the value of $\beta = 0.005$ appeared to be the one that allowed aggregation of faraway areas but without including large cities. $C = 225$ was selected because it is the value as close as possible to 95% of flows with $F_{ij} \geq 11$ without aggregating the EPCI of Marseille-Aix, Toulouse, Bordeaux, Nice, and Montpellier — the big southern cities — into wider zones. Examples of the resulting zoning systems used for the modelling of investment flows are presented on Panel B of Figure 1.

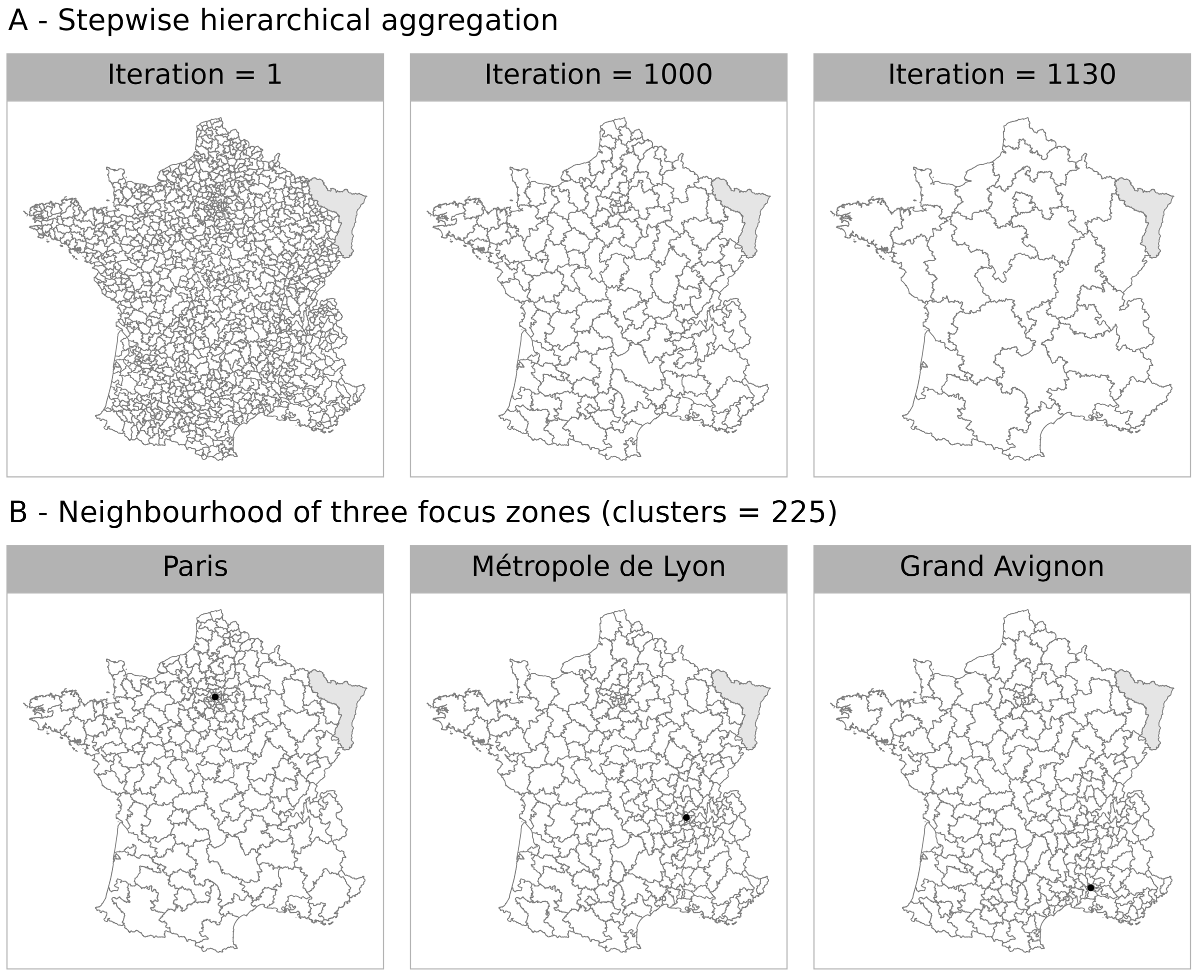
Because we worked with ad-hoc perimeters, a robustness check was performed to assess the impact of the adaptive zoning on the spatial interaction models described in the next section. The results of this robustness check are presented in the Supplementary Material on Figure S2, S3 and in Table S1. They show that the method allows to reduce zero-inflation in the data and to have much more flows and O-D pairs above the confidentiality threshold. These results are obtained with a very limited impact on model calibration.

Based on these new zoning systems, we aggregated buy-to-let investment flows into a

[4]For the special case of Greater Paris, that corresponds to a single giant EPCI, we used the administrative division below, the *Établissements publics territoriaux* (EPT) where "*intra-muros*" Paris is a single unit and municipalities of the *banlieue* are grouped into 11 spatial units.

79 × 225 origin-destination matrix[5].

Figure 1: Steps and example results of the adaptive zoning applied to the EPCI zoning system



4.3 Models and variables selection

Our empirical strategy to test the relative importance of spatial and market factors in explaining the magnitude of investment flows between two territories led us to set up two series of models: a first group of models that only included mass and distance variables, and a second group of models using variables related to house prices as well.

In the first group of models, we first tested a model only taking into account the mass of the origin and the mass of the destination (model I), then we introduced the simplest formulation of the gravity model by adding distance (model II), and to which we added variables on self-containment and contiguity in model III:

[5]For some sparsely populated EPCIs at the periphery of the FUAs, the focus zone was still merged with a contiguous area at the end of the process. In these cases, we forced the disintegration of the two zones, resulting in a zoning system of slightly more areas.

$$F_{ij} = \alpha + \beta_1 \log(\mathrm{population_i}) + \beta_2 \log(\mathrm{population_j}) + \beta_3 \log(\mathrm{distance_{ij}}) + \beta_4 \,\mathrm{internal}_{ij} \\ + \beta_5 \,\mathrm{contiguity}_{ij} + \epsilon_{ij}$$

where F_{ij} is the number of buy-to-let units in j bought by households living in i, $population_i$ is the total number of inhabitants in i, $population_j$ the total number of inhabitants in j, $distance_{ij}$ is the straight line distance[6] between i and j, $internal_{ij}$ is a dummy variable for self-contained flows, $contiguity_{ij}$ is a dummy variable where i and j are contiguous areas, and ϵ_{ij} is an error term.

The second group of models includes variables related to market characteristics. Model IV tests the effect of house prices at the destination, and model V both at the origin and the destination:

$$F_{ij} = \alpha + \beta_1 \log(\mathrm{population_i}) + \beta_2 \log(\mathrm{population_j}) + \beta_3 \log(\mathrm{distance_{ij}}) + \beta_4 \,\mathrm{internal}_{ij} \\ + \beta_5 \,\mathrm{contiguity}_{ij} + \beta_6 \log(\mathrm{median\ price_j}) + \beta_7 \log(\mathrm{median\ price_i}) + \epsilon_{ij}$$

where $median.price_j$ corresponds to the median house price per square metre in j — the place of investment – and $median.price_i$ corresponds to the median house price per square metre in i – the place of residence. We expected the price in the market of investment to have a negative impact of the magnitude of flows. In fact, investors may seek high rental yields that closely depend on housing prices. Conversely, as housing price reflects neighbourhood position, we expected the price in the market of residence to play a positive role.

In order to test whether people tend to invest in markets that are cheaper than the market they live in, we also tested the impact of relative price in model VI:

$$F_{ij} = \alpha + \beta_1 \log(\mathrm{population_i}) + \beta_2 \log(\mathrm{population_j}) + \beta_3 \log(\mathrm{distance_{ij}}) + \beta_4 \,\mathrm{internal}_{ij} \\ + \beta_5 \,\mathrm{contiguity}_{ij} + \beta_6 \,\mathrm{relative.\ price}_{ij} + \epsilon_{ij}$$

where $relative.price_{ij}$ is the ratio of the house prices in j to that in i. This variable was tested separately, because of collinearity with house prices.

In all of these models, the number of investments corresponds to the period 2010–2018.

[6]For the internal distance we applied the same formula as in the initial step of the adaptive zoning.

Data related to population[7] and house prices are for 2014, the middle of the period. The method we used to compute house prices from micro-level transaction data is similar to that of Casanova Enault and Peris (2022) and is described in the Supplementary Material.

Following d'Aubigny et al. (2000), who argued for the use of count models in spatial interaction models, and Burger et al. (2009) who showed the importance of taking into account zero flows in the estimation, we used a Poisson generalized linear model for the gravity equation fitting.

5 Results

5.1 Spatial and territorial dimensions of rental investment

Table 2 shows the results of the spatial interaction models including only mass and distance variables. These variables are all significant with the expected positive or negative signs. In all three models, the masses of the emitting and receiving areas of investments played a positive role. The population of areas had a stronger effect in terms of attractiveness for investment flows than in terms of emissivity. This can be interpreted by assuming that buy-to-let investments target in priority central areas of large cities where rental markets and multiple property ownership are known to be very much developed (André and Meslin, 2021). In addition, such areas in France are concerned by the zoning allowing tax reduction when investing in rental real estate (Le Brun, 2022).

By introducing the distance variable in model II, we could observe a strong drop in the AIC and an increase in the deviance explained[8] by the model (from 0.426 to 0.718). The coefficient associated with distance shows a negative effect (-0.92 ± 0.002), which confirms previous results about distance friction in real estate investment and the fact that private individual investors are home-biased (Levy, 2021; Hochstenbach, 2023). When introducing territorial variables, such as contiguity and self-containment of flows (model III), the likelihood of investors buying rental units in the vicinity of their place of residence is even more manifest. These two variables have the strongest effect on the magnitude of investments (respectively 1.86 ± 0.01 and 4.14 ± 0.01) and considerably improve the goodness-of-fit for the model, as indicated by the AIC and the share of deviance explained by the model (0.879).

The role of geographical factors having been established, it is informative to isolate flows

[7]We used data from the population census carried out by the French statistics office (INSEE).

[8]This measurement is defined as $1 - \frac{D}{D_0}$, where D is the difference of the log-likelihoods between the fitted model and a perfect model and D_0 is the null deviance (the deviance of the model without predictors). This measurement allows quantification of how much the model has improved by adding predictors.

Table 2: Results of spatial interaction models with spatial and territorial variables

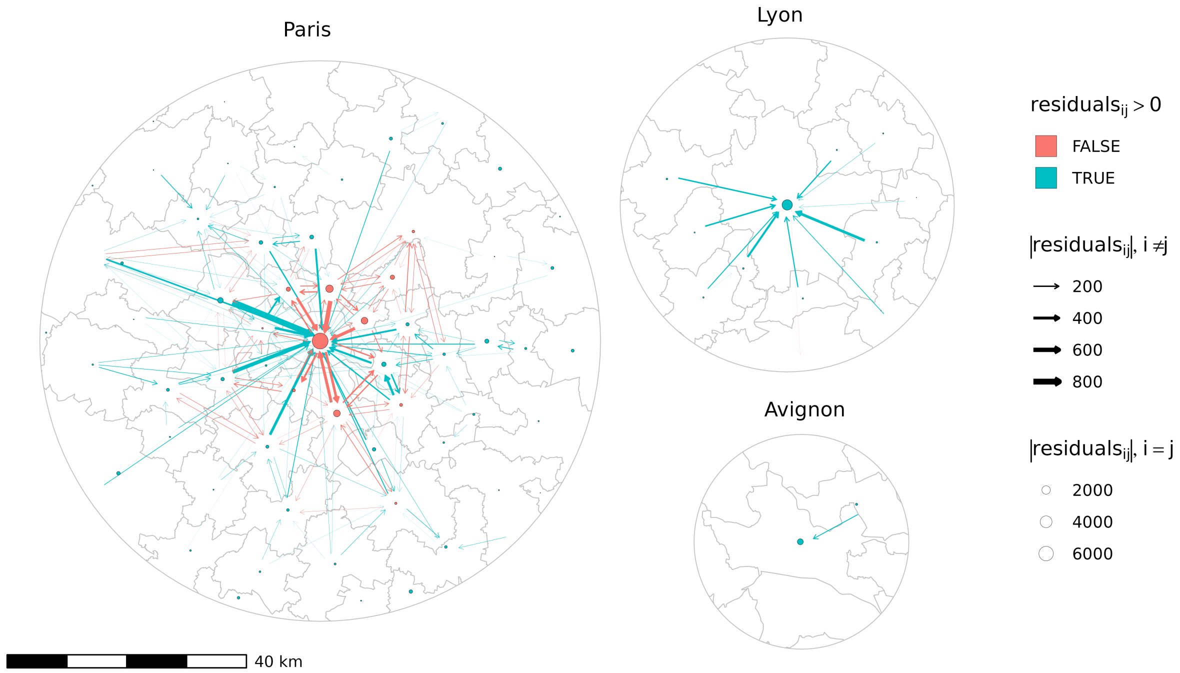
	Dependent variable:		
	F_{ij}		
	(I)	(II)	(III)
$\log(\text{population}_i)$	0.947***	0.782***	0.629***
	(0.002)	(0.002)	(0.002)
$\log(\text{population}_j)$	1.256***	1.039***	0.815***
	(0.003)	(0.002)	(0.003)
$\log(\text{distance}_{ij})$		−0.922***	−0.194***
		(0.002)	(0.003)
internal_{ij}			4.139***
			(0.011)
contiguity_{ij}			1.864***
			(0.010)
constant	−24.787***	−15.775***	−14.822***
	(0.041)	(0.042)	(0.035)
Observations	17,857	17,857	17,857
Log Likelihood	−374,349.900	−193,034.800	−92,807.320
Akaike Inf. Crit.	748,705.800	386,077.700	185,626.600
Explained dev.	0.426	0.718	0.879
Note:			*p<0.1; **p<0.05; ***p<0.01

that are not explained by proximity and population size. Residual analyses are standard in spatial interaction modelling and allow the identification of preferential flows and barrier effects (Guibard, 2021; Pumain, 1986; Rose, 2002). Positive residuals indicate preferential channels of investments between territories, while negative residuals allow barrier effects to be identified. Figures 2 and 3 present the residuals of model III at the local and the national scales.

Two striking tendencies can be observed by zooming in the residuals at the local scale (Figure 2), *i.e.* strong variations in self-contained flows, and the convergence of investments from wealthy peripheral areas to the centres of the agglomerations.

The self-containment of flows was very high in Lyon, especially within the EPCI of Grand Lyon (which echoes the short investment distances observed in the descriptive statistics). These results should be seen in the context of the attractiveness of the Lyon market in recent years, which has experienced some of the highest increases in house prices in France. The underestimation of self-contained flows is also visible in the two territories

Figure 2: Local residuals of model III



Data: Authors' calculations on DGFiP-DGALN-CEREMA Fichiers Fonciers (2009-2020), DV3F (2010-2020)
Note: Only residuals on data points where Fij>10 are displayed

of the FUA of Avignon that were included in the model. In the FUA of Paris, the picture gives much more contrast. The residuals of the model reveal the unequal local geography of income and of house prices, that translates into different purchasing powers regarding geographical sectors. Some territories have positive residuals for self-contained flows, especially the western and some southern suburbs such as Saint Germain Boucle de Seine, Versailles Grand Parc and Paris Est Marne et Bois. These places are wealthy suburbs characterised by high house prices and high median income per capita. Other areas have high negative residuals, especially Paris *intra-muros* as well as several deprived suburbs in the North Est (*i.e.* Plaine Commune and Est Ensemble). These areas are characterized by relatively high house prices and intermediate or low median income per capita. The Paris case is especially striking with local investments being much lower than predicted simply by mass and proximity.

The second main tendency — the convergence of flows from certain suburbs toward the centres of the FUAs — is visible in Lyon and somewhat less so in Avignon. There is a greater contrast in the FUA of Paris. On the one hand, the suburbs of the North East described above as investing little internally are also characterized by negative residuals in the investments towards nearby territories. The strongest negative residuals go from Plaine Commune, the poorest district of Greater Paris, toward Paris *intra-muros*. On the other hand, some suburbs have investment flows to the centre of Paris that are much greater than those predicted by the model. These suburbs are the ones located in the

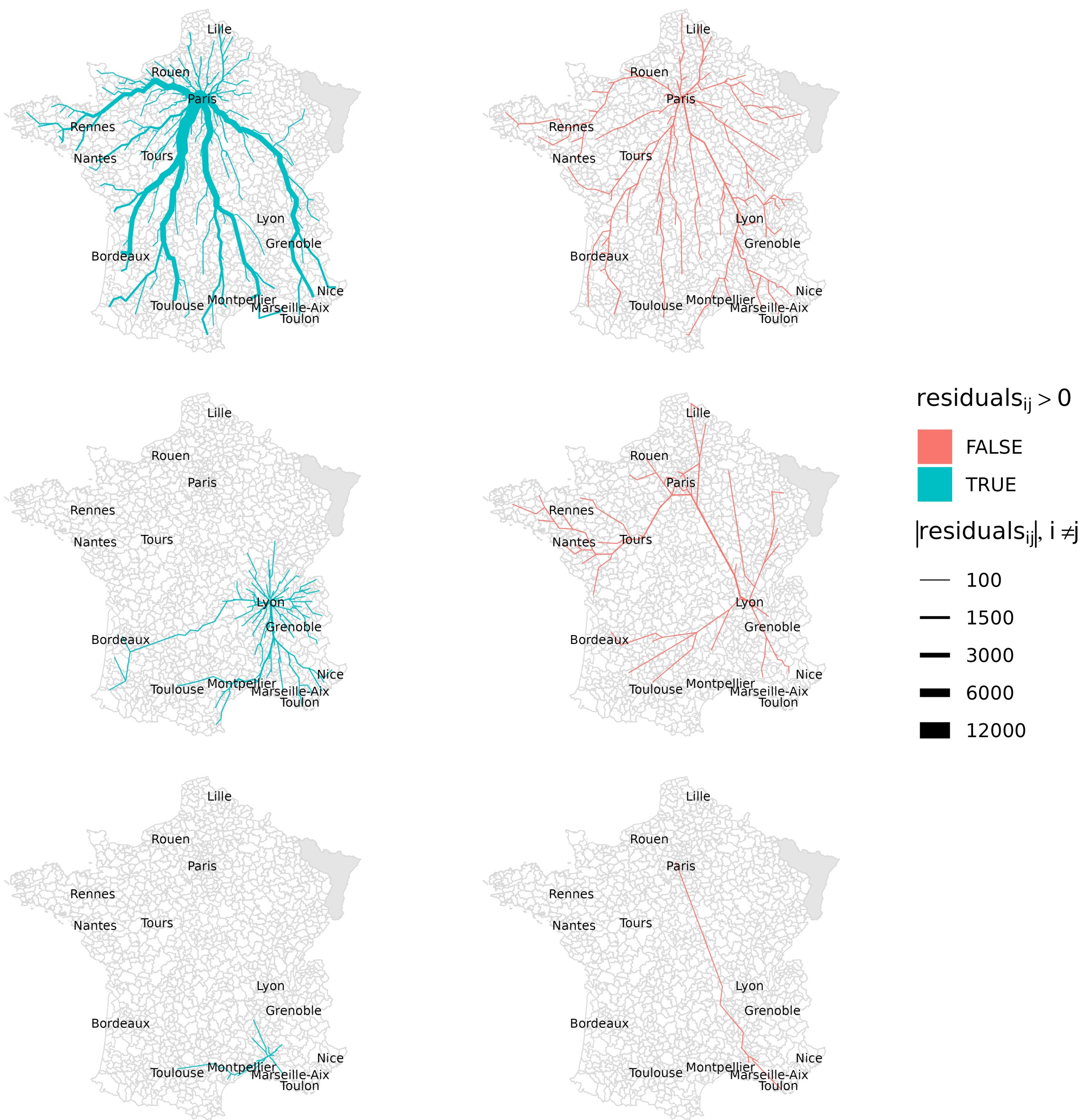
West of the FUA of Paris (Saint-Germain Boucles de Seine, Versailles Grand Parc, Paris Saclay, and Paris Ouest La Défense), as well as one located in the South East (Paris Est Marne et Bois). These EPCIs correspond to the locations of the richest municipalities of the Paris region.

The analysis of residuals between the EPCIs of the three FUAs and distant places (Figure 3) reveals three other striking facts: the regional embedding of investments originating from the FUAs of Avignon and Lyon, the high propensity of households from the FUA of Paris to invest in faraway places, and the importance of long distance investment flowing from the Paris region toward urban areas of the South and the West of France, as well as toward coastal areas.

Positive residuals from Avignon can be observed in nearby big and medium sized cities such as Marseille-Aix, Montpellier, and Nîmes as well as smaller neighbouring towns (i.e. Orange and Carpentras). Negative residuals tend to be in upmarket touristic areas (Luberon, Alpilles, coastal areas of the Var) and in the western part of the FUA of Paris. In the case of Lyon, positive residuals are mostly in the South East quarter of France. They can be observed in nearby small towns (e.g. Villefranche-sur-Saône), and in the nearby medium-sized city of Saint-Étienne as well as in urban areas of the Mediterranean coast (western part of the Côte-d'Azur, Montpellier and Sète). Negative residuals from Lyon can be observed in the northern half of the country (i.e. Rennes, Lille and the suburbs of Paris).

Long distance investments from the Paris region show a very different profile, both in magnitude and geographical patterns. A large share of the positive residuals originates from the EPCIs of Paris Ouest La Défense, Saint-Germain Boucles de Seine, and Paris Est Marne et Bois – the rich suburbs of the FUA – as well as from Paris *intra-muros* where local investments were previously identified as very low. In terms of destination, two cities stand out in the positive residuals originating from the Paris region: Toulouse and Bordeaux. These two metropolitan areas have undergone significant growth over the last two decades, driven in particular by positive net migration among young professionals, students, and managers. Investments from Paris EPCIs are much higher in these large, fast-growing cities located in the South West of France than predicted simply by mass and distance. The other territories characterized by high values of positive residuals are cities along the Mediterranean (Montpellier, Nice, Cannes-Antibes) and the Atlantic coasts (Bayonne-Biarritz-Anglet, La Rochelle and Nantes), as well as the territory including Deauville, Trouville and Honfleur – seaside resorts along the English Channel. The Mediterranean and Atlantic sunbelts, which have recently benefited from amenity migrations in a rather uncontrolled way, appear as attractive areas for investors, even for those who live far away.

Figure 3: Non-local residuals of model III



Data: Authors' calculations on DGFiP-DGALN-CEREMA Fichiers Fonciers (2009-2020), DV3F (2010-2020)
Note: Only residuals on data points where Fij>10 are displayed

Note: For ease of reading, the residual flows from the EPCIs were aggregated at the origin by FUA in order to represent one-to-many flow maps. They were then also grouped with an edge bundling method based on an algorithm using triangulation and approximate Steiner trees (Schoch, 2022). The maps on the left-hand side represent positive residuals and the ones on the right show negative residuals.

5.2 The influence of local market characteristics

The models that include variables on local housing market characteristics are presented in Table 3. In these models, the variables associated with the spatial and territorial dimension of investments remain significant with the expected positive or negative signs.

The percentage of deviance explained by the gravity-market models (models IV, V, and VI) only increased marginally compared to the simple gravity formulation. It means that, at this scale of analysis, proximity remains the main predictor of buy-to-let investment.

Table 3: Results of spatial interaction models with spatial, territorial and market variables

	Dependent variable:		
	F_{ij}		
	(III)	(IV)	(V)
$\log(\text{population}_i)$	0.638***	0.661***	0.697***
	(0.002)	(0.003)	(0.002)
$\log(\text{population}_j)$	0.634***	0.629***	0.709***
	(0.004)	(0.004)	(0.003)
$\log(\text{distance}_{ij})$	−0.044***	−0.043***	−0.113***
	(0.004)	(0.004)	(0.003)
internal_{ij}	4.376***	4.365***	4.324***
	(0.012)	(0.012)	(0.012)
contiguity_{ij}	1.983***	1.973***	1.955***
	(0.010)	(0.010)	(0.010)
$\log(\text{median price}_j)$	0.514***	0.545***	
	(0.008)	(0.009)	
$\log(\text{median price}_i)$		−0.106***	
		(0.009)	
relative price			0.375***
			(0.007)
constant	−17.366***	−16.965***	−15.021***
	(0.053)	(0.062)	(0.035)
Observations	17,857	17,857	17,857
Log Likelihood	−90,849.180	−90,772.050	−91,396.300
Akaike Inf. Crit.	181,712.300	181,560.100	182,806.600
Explained dev.	0.882	0.882	0.881
Note:			*p<0.1; **p<0.05; ***p<0.01

However, the coefficients associated with local housing prices in the investment market are positive both in model IV (0.51 ± 0.01) and model V (0.54 ± 0.01). This means that even when controlling for population size, investments mainly target mostly upmarket areas. Our assumption about investors seeking the high rental yields that low house prices can allow does not hold. Instead, it appears that private individuals seem to favour secure markets. These results are corroborated by the fact that in model VI, the coefficient associated with relative price is positive, at 0.37 ± 0.01. It means that in the case of buy-to-let investments, private individuals tend to favour places that are more expensive than their place of residence. Again, this indicates that investors favour security the over cost-effectiveness and opportunities allowed by cheaper house prices.

Finally, the median house price of the investors' market of residence negatively affects the magnitude of investments originating from them (coefficient of -0.10 ± 0.01). Despite the fact that house prices could partially capture neighbourhood position, it remains clear that, for owner-occupiers, the financial burden associated with living in an expensive area affects their ability to invest in additional housing units.

6 Conclusions

Spatial interaction models using adaptive zoning appear to be an efficient solution for modelling buy-to-let investments occurring at several spatial scales, to uncover preferential channels of investments within and between cities, and to reveal local features.

Although the recent literature has described the emergence of wide ranging investment in the context of financialisation, it seems that space has considerable importance in understanding the mobility of capital related to real estate. Our analysis demonstrates that geographical proximity, approached through different variables, is a strong determinant in explaining the flows of buy-to-let investments by private individuals. This result corroborates other findings that individual landlords tend to invest locally (Hochstenbach, 2023; Levy, 2021), a trend that can also be seen in the strategy of some corporate investors (Coën et al., 2021). Residuals analyses of spatial interaction models also uncover striking facts related to the geography of buy-to-let investment. The first point is that there are important variations in local investments depending on the investor's place of residence. Households from certain areas, such as those from the centre of Lyon, preferentially invest locally, while households from the centre of Paris do not. Second, there is a convergence of investments from certain peripheries of the FUAs – especially the rich suburbs – toward the centre of agglomerations. Finally, we have seen that investors from the Paris FUA are much more inclined to long distance investment. Preferential flows from the FUA of Paris give priority to targeting Toulouse and Bordeaux, two large, fast growing cities in the South West of France, as well as places with residential amenities such as the urban areas of the Mediterranean and the Atlantic coasts.

Regarding market characteristics, we found that the price of housing in the homeowner's market of residence has a negative impact on the magnitude of flows. In fact, a high level of local price acts as an important constraint for any additional investment. Moreover, investors tend to invest in upmarket areas and in places that are more expensive than their market of residence. This finding is consistent with the literature on the increasing assetisation of housing. The majority of multiple property owners are in search of safe investments rather than profitable but risky ones. The housing market appears to be seen less as a means of getting rich quickly and more as a means of securing value in an asset-based welfare context (Benites-Gambirazio and Bonneval, 2022; Doling and Ronald, 2010). Thus, proximity to the place of residence and safety of the investment market are determining factors in housing wealth accumulation by private individuals.

References

Aalbers, Manuel B., Cody Hochstenbach, Jelke Bosma, and Rodrigo Fernandez (2021), "The death and life of private landlordism: How financialized homeownership gave birth to the buy-to-let market". *Housing, Theory and Society* 38.5, pp. 541–563.

Adkins, Lisa, Melinda Cooper, and Martijn Konings (2021), "Class in the 21st century: Asset inflation and the new logic of inequality". *Environment and planning A: economy and space* 53.3, pp. 548–572.

Alstadsæter, Annette, Gabriel Zucman, Bluebery Planterose, and Andreas Økland (2022), *Who owns offshore real estate? evidence from Dubai.* Tech. rep. EU Tax Observatory Working Paper 1.

André, Mathias and Olivier Meslin (2021), "Et pour quelques appartements de plus: Étude de la propriété immobilière des ménages et du profil redistributif de la taxe foncière". *Insee, Document de travail* 2021-04.

Arnold, Céline and Jocelyn Boussard (2017), "L'accès à la propriété en recul depuis la crise de 2008". *Les conditions de logement en France. Insee.*

Arundel, Rowan and Richard Ronald (2021), "The false promise of homeownership: Homeowner societies in an era of declining access and rising inequality". *Urban Studies* 58.6, pp. 1120–1140.

Benites-Gambirazio, Eliza and Loıc Bonneval (2022), "Housing as asset-based welfare. The case of France". *Housing Studies*, pp. 1–15.

Beswick, Joe, Georgia Alexandri, Michael Byrne, Sònia Vives-Miró, Desiree Fields, Stuart Hodkinson, and Michael Janoschka (2016), "Speculating on London's housing future: The rise of global corporate landlords in 'post-crisis' urban landscapes". *City* 20.2, pp. 321–341.

Bonvalet, Catherine and Arnaud Bringé (2013), "Les effets de la politique du logement sur l'évolution du taux de propriétaires en France". *Revue européenne des sciences sociales* 51.1, pp. 153–177.

Burger, Martijn, Frank Van Oort, and Gert-Jan Linders (2009), "On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation". *Spatial economic analysis* 4.2, pp. 167–190.

Cardoso, Rodrigo V and Evert Meijers (2021), "Metropolisation: The winding road toward the citification of the region". *Urban Geography* 42.1, pp. 1–20.

Casanova Enault, Laure, Martin Bocquet, and Guilhem Boulay (2023), "Who owns France? Uncovering the structure of property ownership for a better understanding of the socio-spatial distribution of wealth". *Journal of Urban Affairs*, pp. 1–18.

Casanova Enault, Laure and Antoine Peris (2022), "L'articulation des prix fonciers et immobiliers en France: une géographie des marchés locaux". *Fonciers en débat*, https–fonciers.

Chinco, Alex and Christopher Mayer (2016), "Misinformed speculators and mispricing in the housing market". *The Review of Financial Studies* 29.2, pp. 486–522.

Coën, Alain, Arnaud Simon, and Saadallah Zaiter (2021), "Why is there a Home Bias? An Analysis of US REITs Geographic Concentration 1". *Finance* 42.1, pp. 111–154.

Cottineau, Clémentine, Paul Chapron, Marion Le Texier, and Sébastien Rey-Coyrehourcq (2019), "Incremental Territorial Modeling". *Geographical Modeling*. John Wiley & Sons, Ltd, pp. 95–123.

d'Aubigny, G, Christian Calzada, Claude Grasland, Denis Robert, G Viho, and Jean-Marc Vincent (2000), "Approche poissonnienne des modèles d'interaction spatiale". *Cybergéo* 126.

De Benedictis, Luca and Daria Taglioni (2011), *The gravity model in international trade*. Springer.

Dijkstra, Lewis, Hugo Poelman, and Paolo Veneri (2019), "The EU-OECD definition of a functional urban area". Publisher: OCDE.

Doling, John and Richard Ronald (2010), "Property-based welfare and European homeowners: how would housing perform as a pension?" *Journal of housing and the Built environment* 25, pp. 227–241.

Fernandez, Rodrigo and Manuel B Aalbers (2016), "Financialization and housing: Between globalization and varieties of capitalism". *Competition & change* 20.2, pp. 71–88.

Ferrante, A and R Solotareff (2018), "Entre 1998 et 2015, le patrimoine double, mais diminue pour les 20% les moins dotés". *Les revenus et le patrimoine des ménages*, pp. 27–43.

Fields, Desiree (2017), "Urban struggles with financialization". *Geography Compass* 11.11, e12334.

Fields, Desiree (2022), "Automated landlord: Digital technologies and post-crisis financial accumulation". *Environment and Planning A: Economy and Space* 54.1, pp. 160–181.

Fields, Desiree and Sabina Uffer (2016), "The financialisation of rental housing: A comparative analysis of New York City and Berlin". *Urban studies* 53.7, pp. 1486–1502.

Guibard, Luc (2021), "Déménager en Île-de-France : les ménages aux revenus modestes s'éloignent davantage de Paris". *Les Franciliens - Territoires et modes de vie*. Ed. by L'Institut Paris Région. L'Institut Paris Région, pp. 124–129.

Hagen-Zanker, Alex and Ying Jin (2012), "A new method of adaptive zoning for spatial interaction models". *Geographical Analysis* 44.4, pp. 281–301.

Hagerstrand, Torsten et al. (1968), "Innovation diffusion as a spatial process." *Innovation diffusion as a spatial process.*

Hochstenbach, Cody (2022), "Landlord elites on the Dutch housing market: Private landlordism, class, and social inequality". *Economic Geography* 98.4, pp. 327–354.

Hochstenbach, Cody (2023), "Networked geographies of private landlordism: mapping flows of capital accumulation and rent extraction". *Housing Studies* 0.0.

Hulse, Kath and Margaret Reynolds (2018), “Investification: Financialisation of housing markets and persistence of suburban socio-economic disadvantage”. *Urban Studies* 55.8, pp. 1655–1671.

INSEE (2020), *Tableaux de l’économie française - Edition 2020*. Tech. rep. INSEE.

Josselin, Didier, Delphine Blanke, Mathieu Coulon, Guilhem Boulay, Laure Casanova Enault, Antoine Peris, Pierre Le Brun, and Thibault Lecourt (2023), “Incertitudes liées aux échelles d’estimation des prix immobiliers”. *Imperfection et information géographique 2 - Cas d’études*. Batton-Hubert M., Desjardin E., Pinet F. ISTE éditions.

Kadi, Justin, Cody Hochstenbach, and Christian Lennartz (2020), “Multiple property ownership in times of late homeownership: A new conceptual vocabulary”. *International Journal of Housing Policy* 20.1, pp. 6–24.

Le Brun, Pierre (2022), “Un soutien géographiquement inégal: la sélectivité spatiale des aides publiques à l’investissement immobilier résidentiel des ménages en France”. *Géographie Économie Société* 24.1, pp. 43–68.

Le Goix, Renaud, Laure Casanova Enault, Loıc Bonneval, Thibault Le Corre, Eliza Benites-Gambirazio, Guilhem Boulay, William Kutz, Natacha Aveline-Dubach, Julien Migozzi, and Ronan Ysebaert (2021), “Housing (in) equity and the spatial dynamics of homeownership in France: a research agenda”. *Tijdschrift voor economische en sociale geografie* 112.1, pp. 62–80.

Lei, Rémi, Laure Casanova Enault, Martin Bocquet, and Antoine Peris (2023), *Dynamiques de concentration de la propriété immobilière en France : Apports des Fichiers Fonciers enrichis pour analyser les patrimoines fonciers-immobiliers des individus*. Tech. rep.

Lennartz, Christian, Rowan Arundel, and Richard Ronald (2016), “Younger adults and homeownership in Europe through the global financial crisis”. *Population, Space and Place* 22.8, pp. 823–835.

Levy, Antoine (2021), “Housing Policy with Home-Biased Landlords: Evidence from French Rental Markets”.

McAllister, Pat and Anupam Nanda (2016), “Does real estate defy gravity? An analysis of foreign real estate investment flows”. *Review of International Economics* 24.5, pp. 924–948.

Özogul, Sara and Tuna Tasan-Kok (2020), “One and the same? A systematic literature review of residential property investor types”. *Journal of Planning Literature* 35.4, pp. 475–494.

Paccoud, Antoine (2017), “Buy-to-let gentrification: Extending social change through tenure shifts”. *Environment and Planning A* 49.4, pp. 839–856.

Paris, Chris (2013), “The homes of the super-rich: Multiple residences, hyper-mobility and decoupling of prime residential housing in global cities”. *Geographies of the super-rich*. Edward Elgar Publishing, pp. 94–109.

Peris, Antoine, Evert Meijers, and Maarten Van Ham (2021), "Information diffusion between Dutch cities: Revisiting Zipf and Pred using a computational social science approach". *Computers, Environment and Urban Systems* 85, p. 101565.

Piketty, Thomas (2019), *Capital et idéologie*. Média Diffusion.

Pumain, Denise (1986), "Les migrations interrégionales de 1954 à 1982: directions préférentielles et effets de barrière". *Population (french edition)*, pp. 378–389.

Qiu, Leiju, Yong Tu, and Daxuan Zhao (2020), "Information asymmetry and anchoring in the housing market: a stochastic frontier approach". *Journal of Housing and the Built Environment* 35, pp. 573–591.

Ronald, Richard (2008), *The ideology of home ownership: Homeowner societies and the role of housing*. Springer.

Ronald, Richard and Justin Kadi (2018), "The revival of private landlords in Britain's post-homeownership society". *New Political Economy* 23.6, pp. 786–803.

Rose, Andrew (2002), "Estimating protectionism through residuals from the gravity model". *Background paper for the Fall*.

Rosen, Eva (2014), "Rigging the Rules of the Game: How Landlords Geographically Sort Low–Income Renters". *City & Community* 13.4, pp. 310–340.

Schoch, David (2022), *edgebundle: Algorithms for Bundling Edges in Networks and Visualizing Flow and Metro Maps*. Version 0.4.1.

Shelton, Taylor (2018), "Mapping dispossession: Eviction, foreclosure and the multiple geographies of housing instability in Lexington, Kentucky". *Geoforum* 97, pp. 281–291.

Taşan-Kok, Tuna, Sara Özogul, and Andre Legarza (2021), "After the crisis is before the crisis: Reading property market shifts through Amsterdam's changing landscape of property investors". *European Urban and Regional Studies* 28.4, pp. 375–394.

Vergriete, Patrice (2013), "La ville fiscalisée: politiques d'aide à l'investissement locatif, nouvelle filière de production du logement et recomposition de l'action publique locale en France (1985-2012)". PhD thesis. Université Paris-Est.

Vergriete, Patrice and S. Guerrini (2012), "Stratégies d'investissement locatif et défiscalisation". *Etudes foncières* 158, p. 19.

Wind, Barend, Caroline Dewilde, and John Doling (2020), "Secondary property ownership in Europe: contributing to asset-based welfare strategies and the 'really big trade-off'". *International journal of housing policy* 20.1, pp. 25–52.

Zhang, Yina, Shengjun Zhu, Zihan Yin, Xiaoming Yao, and Zhengyu Wang (2020), "Inter-city housing purchase: a case of China's Yangtze River Delta region". *Cities* 97, p. 102491.

Supplementary Material

Data

We relied on two main sources for tracking rental investments: the Fichiers Fonciers and DV3F databases. The Fichiers Fonciers is a dataset that aggregates information from land registry and local tax services. It provides a description of all the housing units in France as well as the property rights associated with them on the 1st of January of each year. DV3F is an exhaustive dataset of land and real estate transaction deeds.

The data to which we had access cover the Functional Urban Areas (FUAs) of Paris, Lyon and Avignon. The definition used for the FUAs is based on the OECD/EU (Dijkstra et al., 2019). It includes the core city and a wide commuting zone. According to the 2020 population census, the FUA of Paris contains 12.8 million inhabitants and those of Lyon and Avignon respectively 2.1 and 0.3 million inhabitants. We worked on a total of 13.2 million housing units, including all the housing units owned by individuals[9] that are either located in one of the FUAs, or that are located anywhere in France but that belong to someone owning at least one dwelling in one of the FUAs.

Because both datasets are collected for administrative purposes, most of the common categories used in housing studies are not directly available from them. We enhance the transaction dataset with complementary data sources to overcome the lack of category in administrative records.

The first data enhancement task was to create a unique identifier for property owners at the national scale. As the data are collected regionally, property owners have different identifiers depending on the area where they own housing units. For more information on this process, see Lei et al. (2023). Then the category of owner-occupier had to be derived from the occupancy status of the housing unit and by matching the address of the housing unit and the address declared by the owner. Buy-to-let investments were distinguished from other types of acquisitions in a two-step process. First, we matched the newly acquired housing units with information on transactions in order to distinguish purchases from inheritance. Then we looked at the evolution of the occupancy status of the purchased houses two and three years after their acquisition. We kept housing units declared as rented by their owners in the years following their acquisition[10]. As the data we used cover the period 2010–2020, we had to limit the analysis to the period 2010–2018 to observe how the dwellings were used in the two last years of the period. Finally, individuals living at the same address that purchased the same housing unit were grouped into a household, our primary unit of analysis.

[9] Housing units owned by companies or social housing organisations are out of the scope of this research.
[10] This time span is necessary because of the delay in the updating of this information by tax services.

Selection of parameters for the adaptive zoning

Figure S1 presents maps of adaptive zoning with Paris as a focus point. They are the results of the second step of the algorithm presented in the subsection “An adaptive zoning strategy”. The maps all have a fixed number of area ($C = 225$) but the parameter β varies from 0.001 to 0.039. The zoning system that was used in the study is the third one in the first row that corresponds to $\beta = 0.005$.

Figure S1: Neighborhoods for modelling investment flows from Paris with different values of β and $C = 225$

Robustness tests of spatial interaction models using adaptive zoning

In this section, we test the robustness of the parameters of a spatial interaction model using adaptive zoning with different values of β and C. The model specification used in this test is same as model III presented in the subsection “Models and variables selection”. It is a simple gravity model with two additional dummy variables for self-contained flows and interactions between contiguous zones. We test the combination of three β values and three C values (Models 2 to 10). We also compare them with a model without adaptive

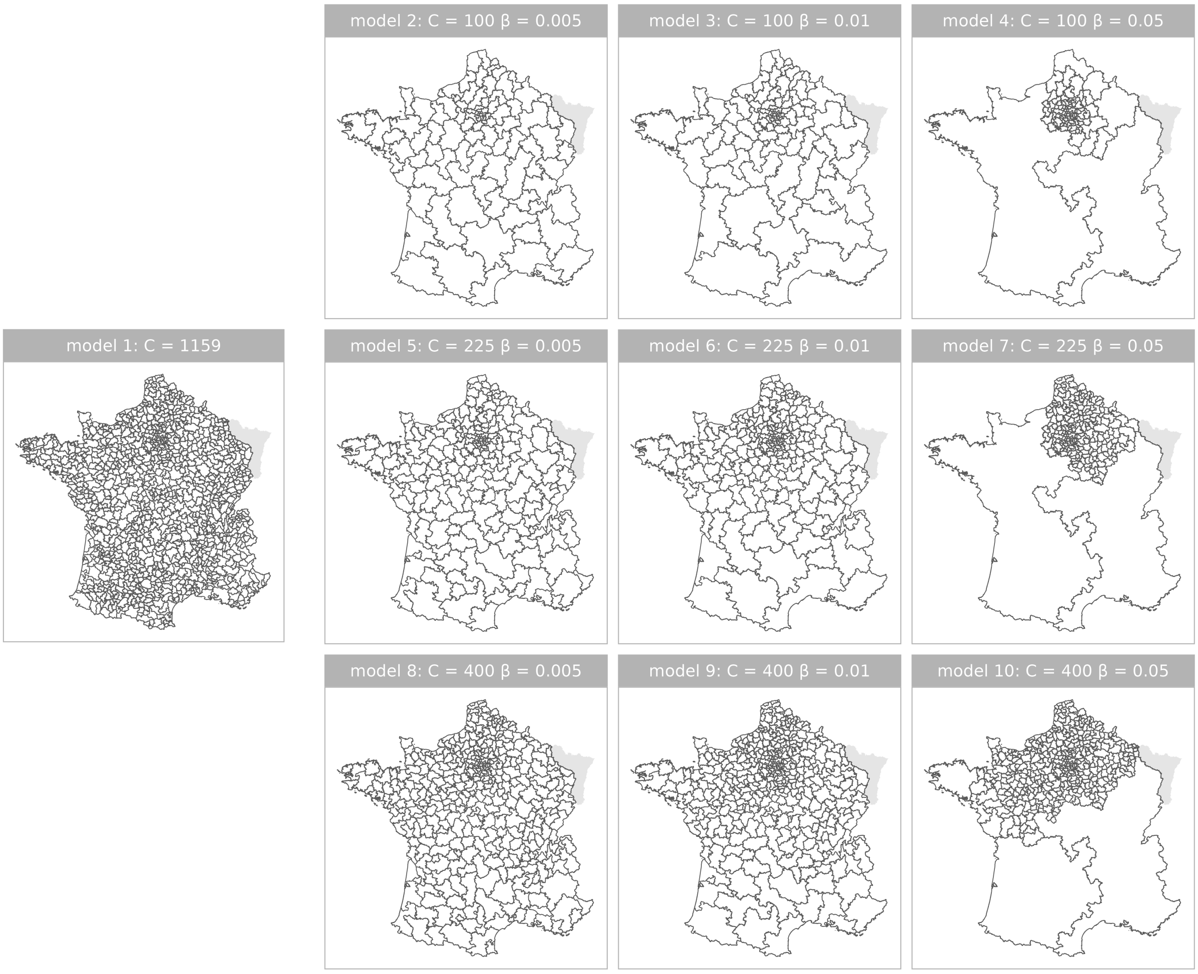


Figure S2: Zoning systems obtained with the combination of β and C values used in the robustness test. Only the ones with Paris as the focus zone are shown.

zoning (Model 1). Figure S2 maps the different zoning systems tested, and Figure S3 shows the values of the parameters obtained for the ten different models. Finally, Table S1 presents model results and goodness-of-fit statistics. It also provides information on the share of O-D pairs with 0 investments, more than 10 investments as well as the total share of investments taken into account in a residual analysis following the rule of $F_{ij} \geq 11$ ("Share of inv.").

Parameters of the gravity models with different zoning systems remain relatively stable. The hierarchy of parameters is the same in all ten models. The parameter associated with mass at the origin ranges between 0.59 and 0.66, and the one associated with mass at the destination ranges between 0.73 and 0.94. Regarding parameters associated with distance and proximity, a similar tendency can be observed (from -0.11 to -0.29 for distance, 1.64 to 1.90 for contiguity and 3.96 to 4.14 for self-containment).

Model 5, with $\beta = 0.005$ and $C = 225$ was selected for the analysis because its parameter values are close to the ones of the model without adaptive zoning and 12.9% of its O-D pairs have more than 10 investments (compared to 2.2% for the model without adaptive

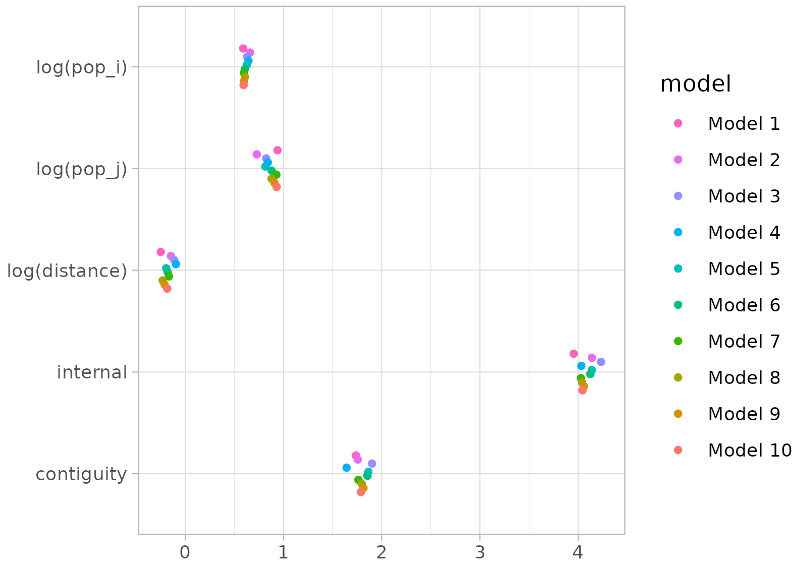


Figure S3: Values of parameters for the ten models

zoning). Moreover, with this zoning system, large cities remain separate entities which is of interest from a thematic point of view.

Computation of price indices

The price index was computed from the DV3F database[11], an exhaustive dataset on land and real estate transactions covering the whole French territory (with the exception of the départements of Alsace and Moselle). We selected all the transactions of apartments and houses between 2013 and 2015. We then filtered out transactions concerning several housing units at the same time. We also applied the filter on unconventional transactions recommended by CEREMA[12], the public organisation maintaining the database. This filter enables the exclusion of sales of rare goods or those with specific sales conditions, which could affect the sale price. It resulted in a total of 1,988,533 transactions. We then computed the price per square meter for each transaction and calculated the median of this value for each territory resulting from the adaptive zoning. Finally, an index on relative prices was also computed for each focus zone, where the price in the market of destination is divided by the price in the market of origin.

[11] https://datafoncier.cerema.fr/dv3f

[12] https://doc-datafoncier.cerema.fr/doc/dv3f/mutation/filtre

Table S1: Results of model III with different values for the number of clusters C and β

C	1159	100			225			400		
β	-	0.005	0.01	0.05	0.005	0.01	0.05	0.005	0.01	0.05
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
(Intercept)	-15.69^{***}	-14.17^{***}	-15.20^{***}	-15.44^{***}	-14.82^{***}	-15.42^{***}	-15.85^{***}	-15.24^{***}	-15.56^{***}	-15.84^{***}
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
log(pop_i)	0.59^{***}	0.66^{***}	0.63^{***}	0.64^{***}	0.63^{***}	0.61^{***}	0.60^{***}	0.61^{***}	0.60^{***}	0.59^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log(pop_j)	0.94^{***}	0.73^{***}	0.82^{***}	0.84^{***}	0.81^{***}	0.88^{***}	0.93^{***}	0.88^{***}	0.91^{***}	0.93^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log(distance)	-0.25^{***}	-0.15^{***}	-0.11^{***}	-0.09^{***}	-0.19^{***}	-0.18^{***}	-0.17^{***}	-0.23^{***}	-0.21^{***}	-0.18^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
internal	3.96^{***}	4.14^{***}	4.23^{***}	4.03^{***}	4.14^{***}	4.12^{***}	4.03^{***}	4.04^{***}	4.06^{***}	4.04^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
contiguity	1.74^{***}	1.76^{***}	1.90^{***}	1.64^{***}	1.86^{***}	1.85^{***}	1.76^{***}	1.80^{***}	1.81^{***}	1.79^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
AIC	268032.58	142836.22	142681.12	119934.54	185626.63	178858.24	137322.25	217673.70	207036.60	169762.99
BIC	268089.13	142878.26	142723.05	119976.39	185673.37	178904.96	137368.97	217723.87	207086.77	169813.15
Log Likelihood	−134010.29	−71412.11	−71334.56	−59961.27	−92807.32	−89423.12	−68655.13	−108830.85	−103512.30	−84875.49
Deviance	212758.81	118603.35	119931.08	104284.85	150164.72	146514.39	117771.97	174302.03	167432.73	140934.64
Num. obs.	91561	8164	8009	7900	17857	17784	17775	31620	31600	31600
Explained dev.	0.883	0.885	0.888	0.920	0.879	0.887	0.927	0.878	0.886	0.916
O-D pairs with $F_{ij} = 0$	78.4%	21.5%	24.1%	46.6%	40.6%	45.6%	67%	55.7%	59.9%	70.8%
O-D pairs with $F_{ij} \geq 11$	2.2%	27.1%	25.1%	16.1%	12.9%	12%	7.3%	7%	6.9%	5.1%
Share of inv.	80.2%	92.6%	93%	95.3%	87.1%	88.6%	93.7%	83.9%	85.7%	90.3%

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: Changes in the number of observations for the same number of clusters are due to the last step of the algorithm that force the disaggregation of the focus zone if it is still part of an aggregate at the last iteration of the neighbourhood creation.