Post-print of: Epifani, Ilenia; Nicolini, Rosella " On the population density distribution across space: a probabilistic approach" in Journal of regional science, Vol. 53 (2013), p. 481-510, ISSN 0022-4146.
The final version is available at DOI: 10.1111/jors.12018

On the population density distribution across space: a probabilistic approach

Ilenia Epifani* and Rosella Nicolini**

*Politecnico di Milano, Dip. di Matematica, P.zza L. da Vinci, 32, I-20133 Milan (Italy).
**Corresponding author: Rosella Nicolini, Universitat Autònoma de Barcelona, Dept. Applied Economics, Edifici B, Campus de la Universitat Autònoma de Barcelona, 08193 Bellaterra (Spain). Ph. +34935814573. *email* <rosella.nicolini@uab.es>.

Abstract

Working within a Bayesian parametric framework, we develop a novel approach to studying the distribution of regional population density across space. By exploiting the Gamma distribution, we are able to introduce heterogeneity across space without incurring an a priori definition of territorial units. Our contribution also permits the inclusion of an approximation of individual preferences as a further driving force in location choices. We perform an empirical application to the case of Massachusetts. Our results demonstrate that a subjective measure of distance performs well in replicating the population distribution across Massachusetts.

Keywords: Agglomerations, Bayesian inference, Distance, Gibbs sampling, Kendall's tau index, Population density.
JEL Classification: C11, C51, R10, R14;

1 Introduction

Explorations of the determinants of territorial organization have flourished in the current economic literature. An understanding of these factors provides new insights that are extremely useful for decision makers in planning interventions that better fit agents' expectations. The complexity of this type of analysis lies in the difficulty of fully grasping the way in which these features combine to produce an evident unequal distribution of the population or activities across space. In this respect, the concept of accessibility turns out to be fundamental in proposing strategies to control for an uneven space structure. The accessibility criterion introduces heterogeneity across space: not all locations can be considered as equivalent from a strictly economic viewpoint. According to the accessibility concept, people enjoy easy access to all of the amenities and other facilities that they desire (Fujita and Thisse, 2002; Song, 1996). They show a particular interest in locating themselves as closely as possible to the place (usually identified as a central business district) that guarantees easy access and proximity to these amenities.

Whenever consumers prefer one location over another, their distribution across space is not expected to be uniform. Several exogenous parametric methods have been proposed to measure geographical accessibility by adopting a sample of selected distance functions but with evident limitations in the generalization of the results. To the same extent, the option of adopting nonparametric or semiparametric estimations needs to fine-tune flexibility in order to capture the territorial endogenous changes of the population density distribution.

In this respect, the analysis we are proposing aims at providing a novel framework for studying population density distribution in a regional setting. The regional dimension embeds an endogenous territorial heterogeneity by accounting for both rural and urban plots. Centering on the case of monocentric distributions, one can provide a general setting for drawing a population density distribution in a dynamic framework in which the size of the regional rural and urban plots may also vary.

Our approach applies to the case of monocentric density. This method of analysis follows the guidelines proposed by Nairn and O'Neill (1988). As we will illustrate in Section 2, the probabilistic approach we are designing provides an interesting framework to overcome the limits associated with the exogenous choice of a distance function. We do not define an a priori structure of the spatial territory, but instead allow our framework to be suitable for adaptation to the largest range of settings. In the same spirit as Helsley and Strange (2007), we also do not define an ad hoc deterministic distance function to represent the accessibility concept. Instead, we model both population density and distance functions as random continuous variables. Our strategy is to conceptualize a representation of a subjective measure of distance embedding subjective preferences (e.g., race, social networks, cultural proximity or similar determinants as discussed in Zenou, 2009) influencing location choices beyond the canonical physical distance. These individual preferences can affect the economic distance between two geographic locations with respect to a common center even if the Euclidean (namely, physical) distance between these locations is identical. By considering distance as a random variable, we can model a broad spectrum of other discriminating factors that make distance a subjective concept for each single individual rather than an objective concept and, therefore, generate spatial heterogeneity. That measure will then be the key tool for understanding the distribution of population density across space. We focus on the population density distribution rather than on the usual conditional expectation to better control for heterogeneity. By taking this approach, we are able to capture the true attractiveness of points in space and are better able to provide point estimations.

In terms of application, our method has been implemented to estimate the regional population density in Massachusetts by exploiting both town and census tract data. To replicate the population density distribution, we construct a setting for which the novelty is to be founded on the association of individual preferences (concerning location decisions) as the main determinant driving population distribution. In this setting, preferences are modeled according to a combination of several factors that picture the attractiveness of a location. The multiple combinations of those factors are the source of an important nonconstant heterogeneity that we must control for in the estimation procedure. Furthermore, our selection also entails the embedding of a degree of spatial dependence into the definition of joint preferences for the location of agents living in the territorial unit that we need to control for. To that end, the Bayesian technique is suitable for this task.[1]

[1] A longer discussion about this issue is included in LeSage and Pace (2009).

From a technical viewpoint, we are able to estimate the determinants of the population density distribution using a Bayesian Gamma model, via a type of Gibbs sampler method. The spatial heterogeneity issue is addressed by accounting for some gamma county random effects (an approach quite similar to the empirical strategy followed by Low and Hining, 2004).

Our technique allows for fitting not only the mean of the true population density distribution, but also its behavior in the upper tail. Indicators of subjective preferences (e.g., income, education, age and ethnical composition) are revealed as the key determinants that allow for the fit of the model to be improved with respect to the baseline specification including physical distance as an exclusive covariate. In addition, the joint magnitude of the subjective-preference variable coefficients is not negligible with respect to that of the physical distance.

The remainder of this paper is organized as follows. Section 2 presents a literature review and describes our analysis strategy. Section 3 develops our analysis setting and illustrates a Gamma-Gamma econometric model with some of its relevant properties. Section 4 applies the model to the Massachusetts case, and Section 5 concludes. All proofs and other materials are included in the Appendix.[2]

2 Literature review and analysis strategy

The study of the sources of heterogeneity in the population density distribution is conducted at two levels: urban and regional. The former environment focuses on the study of the distribution of the population in a city in relation to the changes incurred in its urban structure. The study of the urban population density is an age-old concern in urban economics: population density distribution patterns help us to understand economic and social matters at an urban level. The seminal paper by Clark (1951) is the first attempt to provide a rigorous and formal setting for tackling the previous issues. Since then, interesting empirical methods (grounded to theoretical frameworks) have been developed. In Muth (1969) and Mills (1972), there appears to be a clear concern for estimating and tracking over time and across space changes in population density (at an urban level) and the decline of that density per unit of distance.[3] According to Anas et al. (1998), the first and simplest approach is to work with a monocentric density function. In this case, land use is very simple, and the negative exponential function helps us to explain the strongest empirical regularities. The setting becomes more complex when we refer to polycentric cities that include a central metropolitan area and large subcenters (McMillen, 2004), and some of the attractive characteristics of the metropolitan area still survive in the subcenters.

In contrast, the regional environment addresses the same types of questions in a territory whose distinguishing features are not so uniform as in the previous situation. According to Parr (1985a, 1985b), a region is usually defined as a territory that includes both urban areas and nonmetropolitan areas (usually surrounding the metropolitan areas). Studying the regional population density function is important because the distribution of people is not identical in metropolitan and nonmetropolitan areas. In particular, Parr (1985a, 1985b) states that in nonmetropolitan areas, the density function decreases with distance at a rate different from that found in metropolitan areas. Regardless, both approaches (urban and regional)

[2] An extended version of this study can be found at [http://dx.doi.org/10.2139/ssrn.1873807] with some complementary material.

[3] Other important contributions addressed the problem of obtaining unbiased estimations when working with population density. For example, see McDonald and Bowman (1976), Frankena (1978) and Alperovich (1982), among others.

accept the accessibility concept as a key tool for modeling population distribution density.

In the economic literature, modeling population distribution across space is associated with an investigation of accessibility opportunities. Population density across space is based on the relationship between the value of population density at a given location and the corresponding distance to such a location from a central point, e.g., the central business district (see Nairn and O'Neill, 1988).[4] As listed in Parr et al. (1988) and Song (1996), several possible accessibility functions can be applied, stemming from the variety of distance functions that can be adopted. Commonly, the population density distribution is represented as a discrete function. Hence, it is the product of an index of population density at the center and an accessibility function that decreases as the distance from that center increases. For example, the most commonly applied accessibility function is the inverse of the physical distance from the urban center. Unfortunately, this procedure is very controversial. First, the general distance from a center can be measured in multiple ways. Therefore, we need to identify the most suitable representation of the distance function for the setting we are considering. Second, population location choices can be driven by racial or ethnic preferences, housing issues, demands for local public goods or other factors that cannot always be properly embedded into a simple distance function (for a broader discussion of this issue, see Zenou, 2009 and Glaeser, 2008).

In the regional case, an investigation into the population distribution needs to extend the urban setting to include a further dimension: the heterogeneity of space. A regional space is commonly composed of urban and rural land and, therefore we cannot rely exclusively on the proper modeling technique employed for urban areas. Once more, the problem resides in the definition of the distance function. Parr (1985b) has researched this issue extensively and defines three characteristics that models aiming at replicating the regional population density criteria must fulfill: *i)* the density has to reach a maximum in a short distance from the metropolitan center (which here represents the equivalent of the urban central business district), *ii)* the density should approximate the negative exponential equation between the center and the metropolitan boundary, and *iii)* the function should approximate the Pareto function over the portion from the metropolitan boundary to the regional one. Among the quadratic exponential, linear gamma and log-normal functions that are usually considered as potential candidates, only the log-normal function seems to satisfy these three requirements. However, selecting a log-normal function requires thinking of a specific theoretical setting that endorses its adoption and also allows for a flexible specification that captures regional dynamics across time.

Nairn and O'Neill (1988) go beyond this result: by adopting the techniques of differential equations and asymptotic analysis, they are able to achieve a functional form for the typical behavior of a regional population density that fulfills the three previous criteria. Nevertheless, in their conclusions the authors sketch a possible way of generalizing such a function by conceptualizing a probabilistic density function for which skewness -linked to the standard deviation- is a measure of the dispersion of the population with respect to the distance from the regional center.

Our starting point develops the last argument put forward by Nairn and O'Neill (1988). As a novelty, we consider both population density and the distance function as random continuous variables (rather than deterministic variables). The idea of a random distance function updates the basic notion of (de-

[4] In the urban tradition, Mills and Ping Tan (1980) and McDonald (1989) present interesting surveys of the most common population density function studies in the literature, whereas Bracken and Martin (1988) review the spatial population distribution for census data.

terministic) physical distance by including additional factors that are expected to have an impact on the distance perception of individual agents. In our model, the conditional distribution of the density population (at a given distance) assigns more probability to the larger values of the density population conditioning on lower values of the distance. Our key strategy is to apply the notion of *Negative Regression Dependence* to associate the population density with distance. We measure the Negative Regression Dependence by means of Kendall's tau. This index is an extremely flexible device: it does not change under monotone deterministic transformation of the density population or the distance. Moreover, it can be calculated even if the moments of the density population do not exist. Relying on empirical evidence, we expect that the density is highly skewed against distance and the density population variance increases in the mean. Therefore, we select a Gamma model to fit our goal. In particular, we assume that the population density is Gamma distributed conditionally on distance and negatively regression dependent on distance and that distance is Gamma distributed (Gamma-Gamma model). In this way, we can effectively evaluate to what extent and how rapidly density decreases with distance in terms of the values of the parameters of the Gamma distribution.

Another novelty of this contribution is the exploitation of the Bayesian parametric framework to perform econometric analysis. The Bayesian approach is not commonly used in this type of analysis. As discussed in Anselin (2010), LeSage and Pace (2009) (and previously LeSage, 1997) propose an interesting contribution in terms of the potential applications of the Bayesian approach in spatial econometrics. The Bayesian approach is quite widespread in statistics: it allows for departing from the standard model assumption and controlling for dependence and heterogeneity in the same framework, but from a technical viewpoint, it suffers from a slow convergence rate.

In this respect, as a novelty in the case of spatial analysis, we adopt a Bayesian frailty method. First, a Bayesian frailty trick is a tool to speed up the convergence process. Second, via a Bayesian frailty trick, we are able to control for outliers, distortion of heterogeneity and (potential) spatial dependence within the same statistical framework, without the necessity, for instance, of defining one or more spatial weight matrices. This advantage of the Bayesian approach, in comparison with other methods of estimation, is discussed, for example, in LeSage and Pace (2009). The Bayesian technique provides robust estimations in case of important outliers (which are lessened in their influence), and the introduction of subjective prior information allows for the control of the potential problems of "weak data". In contrast, for instance, the nonparametric methods suffer from problems with nonconstant variance over space and the "weak data" puzzle arises from the adoption of a distance-weighted sub-sample of neighboring data to produce linear regression estimates for each point in space.[5]

Our study concludes with an empirical application of our method to the population density in Massachusetts. This case study represents a suitable situation for replicating the monocentric framework, taking Boston as a neuralgic regional center and characterizing the individual preferences of agents with the common goal of settling close to this center. As also discussed in Glaeser (2008), since Boston is the financial and cultural heart of Massachusetts, it seems quite likely and reasonable that people will display clear preferences for settling in its proximity.[6] Concerning the above discussed advantages of a Bayesian

[5]Quintile regression might be another way to deal with this problem, but additional investigation into the most preferable approach is left for further investigation.

[6]Along this line, we remember the contribution of Rappaport (2009) who showed that the US population has been migrating to places in search of amenities accompanied by increasing wealth.

approach for our Massachusetts sample, in this econometric example we need to control for the distortion of heterogeneity and (potential) spatial dependence: a simple description of the population distribution across Massachusetts exploiting town data confirms the need to treat a clearly uneven territorial distribution (see Figure 4 in Appendix A) with several important density picks.[7] An extremely important population concentration selection appears in the correspondence of the Boston area, and others of a different magnitude are irregularly scattered across the territory.

3 Modeling preferences and probabilistic distribution

Our framework stems from the idea that the spatial distribution of population across a regional territory is the result of the combination of the subjective preferences of residents of that region with respect to the place in which they want to settle. The agents' choice is often based on economic reasons or on the historical and cultural heritage with which all individuals from the same territorial unit agree (Parr, 2007). In a previous stage (not considered here), each citizen defines his or her location choice by maximizing his or her own utility function under a budget and other possible constraints. Hence, once this step is accomplished, we need to define a function representing the distribution of these individuals, taking into account their location preferences. To do so, we identify one-dimensional space (region) with the continuous line $X \in (0, \infty)$ and the total surface of land in each location $x \in X$ equal to one. We consider a continuum N of heterogeneous agents (or households) who choose to reside in that space (Fujita and Thisse, 2002). In this specific framework, we work under the assumption that the agents differ in their degree of preference to settle close to or far from a business center (CBD).[8] In particular, the CBD is defined as a place for which agents either have or do not have an interest. For the sake of simplicity, we focus exclusively on modeling the case in which all individuals aspire to settle as closely as possible to an exogenously selected CBD,[9] and we assume the following individuals agents' preferences.

Assumption 1 *Given two locations $(x, z) \in X$ with $0 < x < z$, then each agent prefers x to z.*

In our setting, the one-dimensional space X is the set of location alternatives allowing for the comparison of two pairs of alternatives, namely, two different location points. According to Mas-Colell et al. (1995), to satisfy a certain degree of consistency in individual agents' choices (here, the individuals' preferences versus their residence lots), we have to assume that their choices satisfy the weak axiom of revealed preferences. Without loss of generality, we define the CBD at zero and X as the distance function from the CBD. In modeling the idea of subjective preferences, the agents' perception of spatial distance must be interpreted as a combination of many different factors: environmental conditions, cultural heritage and so on, beyond the physical distance from the CBD. This way of modeling the distance function allows us to introduce the discriminatory degree of preferences of all the individuals.

To complete the setting, we need to define the way in which we aggregate the individual choices to

[7]The figure of the population density distribution across census tracts is available upon request.

[8]We assume that consumer preferences also encompass easy assess to economic activities as well as employment centers. Because we do not control for them explicitly, they will contribute to the gamma random effects in the model described in Section 4.1.

[9]This framework could be symmetrically adapted to the case in which an individual seeks to settle far from a specific exogenous place. Instead, problems may arise when the variance across their preferences is extremely large. For a discussion about this issue see Fujita and Thisse (2002).

shape the population distribution. Let Y denote the population density in space and let $F_{Y|X}$ be its cumulative distribution function conditional to the distance X: $F_{Y|X}(y \mid x) = P(Y \leq y \mid X = x)$. Assume the following.

Assumption 2 *Y is* negatively regression dependent on *X, i.e.,*

$$F_{Y|X}(y \mid x_1) \leq F_{Y|X}(y \mid x_2),\ \forall y \in \mathbb{R} \text{ and } \forall x_1 < x_2 \ . \tag{1}$$

Inequality (1) says that the proportion of territorial units at a distance x_1 from the CBD with a population density at most equal to y is no greater than the proportion of the more distant territorial units (x_2) with a population density at most equal to y. According to Assumption 2, large distances X from the CBD tend to be associated with small densities of population Y. The hypothesis we introduce regarding agents' preferences implies that the population density distribution is likely to decrease as the distance from the CBD increases. Thus, Assumption 2 is the translation in a probabilistic setting of Assumption 1, representing the idea that all citizens prefer to settle close to the CBD.

Negative regression dependence is an asymmetric concept: the fact that Y is negatively regression dependent on X does not imply that X is negatively regression dependent on Y. Furthermore, if Y is negatively regression dependent on X, then $g(Y)$ is negatively regression dependent on $h(X)$, for every increasing functions g, h. In other words, negative regression dependence is robust with respect to every increasing deterministic transformations of X or Y or X and Y. Finally, the fact that Y is or is not negatively regression dependent on X is a nonparametric question, in the sense that the concept of negative regression dependence applies whatever the parametric assumptions about the distribution of X, Y. [10]

A nonparametric way of measuring the degree of the negative regression association of Y on X is given by Kendall's tau.

Definition 3 Kendall's tau τ *of a random couple* (X, Y) *is*

$$\tau = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0),$$

where (X_1, Y_1), (X_2, Y_2) *are two independent copies of* (X, Y)*. If the random couple* (X, Y) *is continuous, then* $\tau = 1 - 2\pi_d$*, where* $\pi_d = P((X_1 - X_2)(Y_1 - Y_2) < 0)$.

Kendall's τ index assumes values in interval $[-1, 1]$ and is negative if and only if $\pi_d > 1/2$. If τ is negative, random variables X, Y are called *discordant or negatively associated*: τ is negative (if and) only if the probability that either $x_1 < x_2$ is associated with $y_2 > y_1$ or $x_1 > x_2$ with $y_2 < y_1$ is greater than $1/2$, meaning that X and Y are discordant if their values are dissociated with a high probability, namely

[10] From a historical point of view, the concept of negative regression dependence was introduced by Lehmann (1966). More recently, the "positive" regression dependence has been used in insurance theory to model the dependence between the insurable and uninsurable risk (Dana and Scarsini 2005). Moreover, the positive regression dependence is a notion similar to that of "affiliation" discussed in Milgrom and Weber (1982) and used in auction theory. As demonstrated in Dana and Scarsini 2005, affiliation implies positive regression dependence.

From an economic viewpoint, the concept of *Negative Regression Dependence* identifies with an operative device to fully shape the preferences of individuals when referring to their location choices inside a regional territory.

greater than one-half. On the other hand, we prove in Appendix B that *if Assumption 2 is true, then X, Y are discordant, i.e., $\tau < 0$* .

Remark 4 Similar to Negative Regression Dependence, the value of Kendall's τ of a couple (X, Y) also does not change under every increasing monotone deterministic transformation of X or Y or X and Y. In fact, $P(X_2 > X_1, Y_2 < Y_1) = P(g(X_2) > g(X_1), h(Y_2) < h(Y_1))$, for any increasing function g, h. This remark turns out to be very useful for two reasons. First, it simplifies the computation of Kendall's τ. Second it sheds light on what parameters effectively determine Kendall's τ and thus influence the dependence of the density population Y from the distance to CBD X, once a joint model for X, Y is assumed.

By referring to Kendall's τ, we are able to provide a measure of agents' preferences in terms of location choices.[11] Therefore, the use of Kendall's τ allows us to discriminate between the distribution functions that can be suitable for our framework of analysis and all of the other distribution functions.

How much X, Y are discordant depends on the joint distribution of distance and population density. There is no one way to define a priori population density and spatial distance distributions: any positive random variables for which Assumption 2 holds can be selected. To provide the flavor of this selection technique, we focus on a couple of functions that are suitable for representing the population density distribution across a regional space. The following Log-normal and Gamma models have already been discussed in Parr at al. (1988) and Song (1996).

Example 5 (Log-normal Model) Let Y be defined as

$$\ln Y = \alpha_0 - \alpha X + \epsilon. \tag{2}$$

where ϵ is a random disturbance term with zero mean and constant variance and X and ϵ are independent. Then, Y is negatively regression dependent on X as $\alpha > 0$. In fact, the conditional cumulative distribution function of Y given $X = x$ corresponds to that of $\exp\{\alpha_0 - \alpha x + \epsilon\}$, which clearly stochastically decreases in x if $\alpha > 0$. In Equation (2) the density of population Y is modeled as a negative exponential function of the distance from a territorial center (e.g., a CBD), and the parameter α is the density gradient that describes how rapidly the density decreases with distance. This corresponds to a classical analysis of the accessibility problem where, by assumption, α is assumed to be greater than zero. Refer, for example, to the estimation function of the accessibility measure numbered one in the first row of Table 1 in Song (1996). Thus, this example emphasizes that the classical log-normal regression analysis of the location accessibility satisfies Assumption 2.

Example 6 (Gamma-Gamma Model) Let $\eta(x) : (0, \infty) \mapsto (0, \infty)$ be a monotone increasing function and suppose that conditional on $X = x$, Y has Gamma distribution[12] with shape θ and rate function

[11] In the spirit of Parr (1985a), we are interested in working within a framework that is suitable for shaping the population density distribution for territories including both urban and rural areas. This reason explains the decision to focus on the regional rather than the urban dimension.

[12] A Gamma probability density with shape α and rate β has positive support, and its probability density on $(0, \infty)$ is

$$\frac{\beta^\alpha x^{\alpha-1} \mathrm{e}^{-\beta x}}{\Gamma(\alpha)},$$

$\theta \cdot \eta(x)$:

$$Y|X = x \sim \Gamma\left(\theta, \theta\eta(x)\right) \ . \tag{3}$$

Thus, Assumption 2 is satisfied. For technical details, see Appendix C. The conditional mean of Y given $X = x$ is $\mathrm{E}(Y|X) = 1/\eta(X)$, and the conditional coefficient of variation CV is constant and

$$CV = \sqrt{\mathrm{Var}(Y|X)}/\,\mathrm{E}(Y|X) = \sqrt{\mathrm{E}^2(Y|X)/\theta}/\,\mathrm{E}(Y|X) = \frac{1}{\sqrt{\theta}} \ .$$

If $X \sim \Gamma(\alpha, \beta)$ with $\alpha, \beta > 0$, we shall denote the joint model: $Y|X \sim \Gamma\left(\theta, \theta\eta(X)\right)$ and $X \sim \Gamma(\alpha, \beta)$ as $\Gamma(\theta, \theta\eta(X)) \times \Gamma(\alpha, \beta)$. Two interesting models correspond to the following alternative choices for the link function $\eta(x)$: $\eta(x) = ax + b,\ a > 0$ and

$$\eta(x) = \frac{\mathrm{e}^{aX}}{b} \quad a, b > 0. \tag{4}$$

In light of Remark 4 and of the properties of the Gamma distributions family, we have that the models $\Gamma(\theta, \theta aX) \times \Gamma(\alpha, \beta)$ and $\Gamma(\theta, \theta X) \times \Gamma(\alpha, 1)$ share the same Kendall's τ.[13] It then follows that the Kendall's τ of a Gamma-Gamma model $\Gamma(\theta, \theta aX) \times \Gamma(\alpha, \beta)$ depends only on the shape parameters α and θ and is independent of both of the two rate parameters a and β.

For $\eta(x) = ax + b$, Gamma-Gamma models $\Gamma(\theta, \theta(aX + b)) \times \Gamma(\alpha, \beta)$ and $\Gamma\left(\theta, \theta(a/\beta X + b)\right) \times \Gamma(\alpha, 1)$ share the same τ. In other words, if $b \neq 0$ then τ depends on the shape parameters θ, α and the gradient a/β computed as a pure number. Without loss of generality, in our numerical calculations of τ for a Gamma-Gamma model with $\eta(x) = ax + b$, we can let $\beta = 1$.

If $\eta(x)$ is an exponential function as in (4), then

$$\tau(X, Y) = \tau\left(\beta X, \frac{Y}{b}\right) = \tau(W, Z), \qquad \text{where } W \sim \Gamma(\alpha, 1),\ Z|W \sim \Gamma(\theta, \theta\mathrm{e}^{aW/\beta}) \ .$$

Additionally, in this case, τ depends only on the coefficient of variation $\mathrm{CV}(Y|X) = \theta^{-1/2}$, on the shape α of X and on the ratio a/β. As a consequence, in our numerical calculations of τ for a Gamma-Gamma model with link function (4), we can set $\beta = 1$.

As a matter of fact, the expression for the Kendall's τ of a Gamma-Gamma model can only be evaluated numerically or via simulation. Some simplifications for τ arise when α and θ are integers. For example, when $\eta(x) = ax$ and $\alpha = \theta = 1$, one has

$$\tau(X, Y) = 2\int_0^\infty \int_{x_1}^\infty \int_0^\infty \int_{y_2}^\infty x_1 x_2 \mathrm{e}^{-x_1 y_1 - x_1 - x_2 y_2 - x_2}\ dy_1 dy_2 dx_2 dx_1 - 1 = -0.5 \ .$$

where $\Gamma(\alpha)$ represents the Gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\mathrm{e}^{-\beta x}\ dx$.

[13]In fact, if $W = \beta X$ and $Z = mY$ with $l, m > 0$, then $f_W(w) \sim \Gamma(\alpha, 1)$ and

$$f_{Z|W}(z|w) = \frac{1}{m} \times f_{Y|X}\left(\frac{z}{m}|\frac{w}{\beta}\right) = \frac{\left(\theta\eta\left(\frac{w}{\beta}\right)/m\right)^\theta}{\Gamma(\theta)} z^{\theta-1} \exp\Big\{-z\frac{\theta}{m}\eta\left(\frac{w}{\beta}\right)\Big\} \ .$$

Therefore, the Kendall's τ of a $\Gamma(\theta, \theta\eta(X)) \times \Gamma(\alpha, \beta)$ model is equal to the Kendall's τ of a $\Gamma(\theta, \theta\eta(X/\beta)/m) \times \Gamma(\alpha, 1)$ model, and for $m = a/\beta$ we provide the result.

In general, to evaluate τ we simulated $N = 20\ 000$ independent random couples $\{(X_i, Y_i)\}_{i=1}^N$ from the distribution of (X, Y) and calculated the *empirical Kendall's coefficient of concordance*:

$$R_K = \frac{C - D}{N(N-1)} \ ,$$

where C is the number of "concordant" couples and D the number of "discordant" couples.[14]

Table 1(a) shows the results of the simulations for the linear link function $\eta(x) = ax + b$ and Table 1(b) for the exponential $\eta(x)$ in (4). One observes in Table 1(a) the following: *i)* for fixed b and θ, the negative

(a) The case $\mathrm{E}(Y|X) = 1/(aX + b)$.

			α			
a	b	$\mathrm{CV}(= 1/\sqrt{\theta})$	1	10	50	100
		0.2	-0.889	-0.646	-0.392	-0.289
0.5	0	1.0	-0.499	-0.175	-0.079	-0.063
		4.0	-0.081	-0.023	-0.011	-0.009
		0.2	-0.582	-0.580	-0.385	-0.292
0.5	1.0	1.0	-0.147	-0.148	-0.078	-0.067
		4.0	-0.021	-0.019	-0.006	-0.013
		0.2	-0.127	-0.308	-0.299	-0.247
0.5	10.0	1.0	-0.018	-0.057	-0.063	-0.058
		4.0	0.000	-0.009	-0.006	-0.006
		0.2	-0.888	-0.648	-0.381	-0.297
1.0	0	1.0	-0.504	-0.175	-0.071	-0.058
		4.0	-0.078	-0.019	-0.013	0.002
		0.2	-0.704	-0.614	-0.382	-0.294
1.0	1.0	1.0	-0.227	-0.160	-0.074	-0.058
		4.0	-0.023	-0.011	-0.011	-0.010
		0.2	-0.237	-0.413	-0.337	-0.264
1.0	10.0	1.0	-0.047	-0.098	-0.065	-0.052
		4.0	-0.008	-0.005	-0.006	-0.005
		0.2	-0.888	-0.647	-0.396	-0.289
10.0	0	1.0	-0.507	-0.174	-0.072	-0.052
		4.0	-0.077	-0.023	-0.004	-0.005
		0.2	-0.866	-0.643	-0.387	-0.295
10.0	1.0	1.0	-0.433	-0.175	-0.072	-0.058
		4.0	-0.049	-0.015	-0.006	-0.005
		0.2	-0.709	-0.614	-0.387	-0.297
10.0	10.0	1.0	-0.233	-0.157	-0.072	−0.053
		4.0	-0.029	-0.021	-0.011	-0.010

(b) The case $\mathrm{E}(Y|X) = e^{-aX}$.

		α			
a	$\mathrm{CV}(= 1/\sqrt{\theta})$	1	10	50	100
	0.2	-0.670	-0.916	-0.963	-0.974
0.5	1.0	-0.233	-0.579	-0.788	-0.846
	4.0	-0.026	-0.105	-0.205	-0.266
	0.2	-0.810	-0.958	-0.981	-0.987
1.0	1.0	-0.388	-0.759	-0.890	-0.923
	4.0	-0.056	-0.177	-0.356	-0.436
	0.2	-0.978	-0.996	-0.998	-0.028
10.0	1.0	-0.878	-0.975	-0.988	-0.041
	4.0	-0.387	-0.737	-0.873	-0.030

Table 1: Kendall's τ for the Gamma-Gamma model in Example 6. Table (a) refers to the case $\mathrm{E}(Y|X) = 1/(aX + b)$ with coefficient of variation $\mathrm{CV} = 1/\sqrt{\theta}$ and $\beta = 1$. The variability in Kendall's τ when $b = 0$ for different values of a -for example, $\tau = -0.499, -0.504, -0.507$, for $\theta = 1$ and $a = 0.5, 1, 10$, respectively- is exclusively due to the simulation errors. Table (b) refers to the case $\mathrm{E}(Y|X) = e^{-aX}$ with $\beta = 1$.

dependence decreases as α increases and the Kendall's τ approaches zero from below as $\alpha \to \infty$; *ii)* given the shape parameter α of the Gamma distance X, the negative dependence decreases with the conditional coefficient of variation $\mathrm{CV} = 1/\sqrt{\theta}$ of Y given X. Larger values of CV imply a larger dispersion for Y, whereas larger values of α imply a larger variance for X and, thus, a smaller dependence between X

[14] Two pairs are *concordant* if both members of a couple are larger than the respective members of the other couple, whereas two pairs are *discordant* if the two members of one couple differ in the opposite sense from the respective members of the other couple.

and Y in both cases. Finally, given θ, $b(\neq 0)$ and α, the dependence between X and Y increases as the gradient a increases and is more relevant with higher values of b and θ. As shown in Table 1(b), the same findings hold in the case of a Gamma-Gamma model with the exponential link function $\eta(x)$ in (4).

Measuring the degree of dissociation in the location choices of agents by means of τ, we note that the negative dependence proportionally reinforces the shape θ, i.e., when the distribution density polarizes. In this sense, the parameters of the function replicate the degree of skew of the density and allow the replication of the different gradients of the density across space.

Taking into consideration these land dynamics, whenever the territorial center of reference enlarges, the preferences of agents coincide because they all want to settle close to this center, and this situation occurs for any distance function shaping the space. From a regional perspective, the enlargement of the CBD results in more blurred agents' preferences with respect to a possible location close to the CBD, especially when the shape parameter of the distance function α increases. This result is due to the polarization of the distance function that makes this spatial dimension disappear. Referring to some empirical evidence (e.g., Chicago as in McMillen 2004), the enlargement of the central business district corresponds to the creation of subcenters (which may be driven by some external factors such as firm location) surrounding the central metropolitan area. The new subcenters reinforce the size and attractiveness of the urban area (namely, the CBD in our framework) with respect to the rural area. Therefore, from a regional perspective, an extreme case would be to obtain an ever-enlarging center to the point where the regional space is reduced to a single urban point without a rural plot.

A model like the Gamma-Gamma one -with constant coefficient of variation- is very useful if we expect the variance of Y to increase with its mean. If $\eta(x) = \mathrm{e}^{aX}/b$, the conditional mean $\mathrm{E}(Y|X = x)$ coincides with the notion of accessibility, as does the distance from a center. Once more, the parameter a is a measure of the density gradient describing the decreasing speed of the density with respect to the distance, whereas the parameter $\ln b$ describes the density at or near the CBD. In other words, the negative dependence of the density population on the distance from the CBD does not depend on b, as we observed earlier. Nevertheless, if we take a log link, i.e., $\ln \mathrm{E}(Y|X) = \ln b - aX$, then Model (3)-(4) can be analyzed under the generalized linear model setup (see McCullagh and Nelder, 1989).

Both the Log-normal model at (2) and the Gamma model in (3)-(4) have a constant coefficient of variation, and thus both are often useful in similar problems. Using the Gamma model (3)-(4) to address accessibility is equivalent to adopting a regression model with multiplicative gamma errors for the original data, whereas Log-normal model (2) provides an additive regression model for the logarithms of density. In other words, in a Gamma model, the density population is measured on the original scale, whereas in a Log-normal model, the scale of measurement of the density is logarithmic. McCullagh and Nelder (1989) suggest the assumption of a Gamma distribution if it is preferable to work with data in original scale (an issue also discussed in Parr, 1985a). Such is the case if, with data on the spectrum $Y_1, \ldots, Y_n$, the sum $Y_1 + \cdots + Y_n$ has a physical meaning. In our case study in Section 4, the sum of population densities $Y_1, \ldots, Y_n$ of all the n territorial units of space (alternately, census tracts and towns in Massachusetts) is the density population of that space. Furthermore, Firth (1988) proved that, under reciprocal misspecification, a Gamma model performs slightly better: analyzing log-normal data with a gamma model is more "efficient" than analyzing gamma data assuming log-normality.

	Town		Census tract	
	pD	*DIC*	pD	*DIC*
Boston	21.2	**461.5*****	19.3	**7043.6*****
Greenfield	17.5	536.5	34.0	7162.8
Brockton	19.9	486.9	19.6	7130.8
Lowell	16.8	523.9	19.5	7044.7
New Bedford	19.9	522.2	23.6	7148.9
Pittsfield	23.0	532.8	31.0	7166.2
Springfield	16.6	538.3	27.3	7165.1
Lynn	22.1	467.3	18.5	7162.3
Amherst	17.6	537.7	35.1	7163.1
Edgartown	24.9	529.7	29.1	7155.5
Worcester	15.8	534.4	21.5	7170.4
Quincy	21.0	462.7	19.2	7050.9
Barnstable	31.8	523.7	29.3	7147.7
Nantucket	27.2	528.9	36.6	7162.0

Table 2: Center selection

4 A case study for the Gamma-Gamma Model: the population density in Massachusetts

To illustrate the properties of our framework, we propose an empirical application to the case of Massachusetts. Data are taken from the US Census Bureau and refer to the year 2000. We alternately consider census tract and town data to assess the robustness of our novel approach, especially when increasing the sample size to encompass a larger degree of spatial variation. To quantify the available information, the database dealing with census tract territorial units includes 1,361 entries, whereas the town data focus on 351 towns belonging to the state. They are both grouped by 14 counties.

In our monocentric theoretical framework, we first tackle the issue of identifying the CBD. To this end, we proceed by selecting the 14 county seats in Massachusetts as the most representative centers and suitable to be considered as preferred locations. We compute the bilateral distances from *i)* any central points of the 1,361 census tracts and *ii)* any town between the 351 towns in our sample.[15] We then consider (in alternation) each of 14 (Euclidean) distances as the unique explanatory variable and run a reduced form of our regression Model (9), which we label *Model 0*. We perform the estimations using the set of census tract and town data. Of the 14 models, we select the "best match" of the distribution of the population density, according to the Bayesian Deviance Information Criterion (DIC): the model yielding the smallest value of DIC is chosen.[16] We summarize the DIC results obtained for CBD selection in Table 2, from which we deduce that the distance from Boston ensures the best fit to the real data.[17] In light of the current literature, this result is not totally unexpected. Empirical evidence does suggest that the geographic center of reference in Massachusetts is the capital city of Boston. Parr (2007) asserts that Boston can be considered as an example of a "Built city" because of its relative importance for

[15]These distance values have been computed by using the ArcView GIS program. In performing the selection of the census tract of the 14 county seats, we follow the methodology proposed by McDonald (1987).

[16]The DIC is intended as a generalization of Akaike's Information Criterion (AIC); DIC is given by the deviance (i.e., minus two times the likelihood) calculated at the mean of the posterior distribution of all parameters plus two times the effective numbers of parameters: pD. Models with a lower DIC are preferred over those with a larger DIC.

[17]If we accounted for models whose differences in DIC with respect to the model with the lowest DIC are less than five units, other potential CBDs would have to be taken into consideration. This is the case with Quincy for town data and with Lowell for census tract data. We performed estimations with Quincy and Lowell as the CBDs and the results are very similar to the Boston ones (in particular, with respect to the significance of the covariates), but with a slight lower fit. These results are available upon request.

Massachusetts as a focus of economic activity and a demand for labor. Therefore, we choose Boston as our center for carrying out our empirical analysis.

In Table 5 in Appendix A we elaborate some descriptive statistics at the town level. First, we recognize in Table 5 a negative association between the population density and distance from Boston: a lower population density is associated with a greater distance from Boston. Here the distance is understood as the shortest distance from each county seat to Boston.

Nevertheless, the simple physical distance from a selected (attractive) CBD is not the unique factor shaping the population distribution across space. Other factors contribute in a sensitive way to define the preference of citizens to settle in specific locations. In this respect, Wood and Parr (2005) have already introduced the concept of *transaction space* in which they discuss how space can influence the coordination issues among agents (or firms) in the presence of a spatial differentiation in institutional, cultural or language characteristics. Hence, cultural proximity or infrastructure quality can affect the economic distance between two places, even if the Euclidean distance between them is identical. In our theoretical framework, we aim to recover this wider concept, which is the reason why we seek to identify a number of predictors of population density in addition to physical distance. We argue that population location preferences are likely to include a number of environmental characteristics as a discriminating component of the agents' preferences in terms of location choices. We then augment the traditional monocentric regional density model (as in Beckman, 1971) centered on the decay effect of distance ("$Dist$") by adding the ethnic ("Mix"), age ("Age") and education ("$Educ$") composition as well as the income ("Inc") level of the population living in each territorial unit to which we refer. Our selection relies on the main outcomes of the social network literature in spatial economics (as discussed in Zenou, 2009 and Helsely and Zenou, 2011): individuals often demonstrate a tendency to cluster with other people who share, for instance, the same culture or have a similar income level.

Furthermore, the econometric model we introduce includes some unobserved state variables $X_1, \ldots, X_{14}$ one for each county to capture the relationship between the census tracts (or towns) of the same county and the heterogeneity among the counties. In some sense, X_i summarizes all the predictors of the population density, either unobservable or observable but neglected, common to all the census tracts (towns) in the ith county.

With regard to the probabilistic assumption on the law of the population density Y, looking at the empirical means and the standard deviations of the population density (and income) across counties of Massachusetts in Table 5 in Appendix A, we note an increase in the standard deviation with larger mean values. This feature of the data makes a Gamma model a good candidate for adoption because it is characterized by a constant coefficient of variation. The next section describes the proposed econometric two-stage Gamma-Gamma model.

4.1 The Gamma-Gamma Model

Let Y_{ij} be the density of the population of the jth census tract (or, respectively, town) within the ith county, D_{ij} its distance from the selected CBD (here, Boston), $V_{ij}^{(s)}$ the local covariates named in Table 3,

$V^{(2)}$	$V^{(3)}$	$V^{(4)}$	$V^{(5)}$
Mix	*Inc*	*Educ*	*Age*

Table 3: Legenda of local covariates in Model (5)

and Z_i the size of free land (namely, the water areas) in county i. We propose the following model:

$$
\begin{aligned}
&Y_{ij}|X_i \sim \text{Gamma}\left(\theta, \theta X_i e^{-\left(a_i D_{ij} + \sum_{s=2}^{5} \beta_s V_{ij}^{(s)} + \sum_{s=3}^{5} \beta_{3+s} V_{ij}^{(s)} D_{ij}\right)}\right) \quad \text{independent for all } j = 1, \ldots, n_i \\
&X_i \sim \text{Gamma}(\alpha, \alpha B_i) \quad \text{independent for all } i = 1, \ldots, k \\
&a_i = \beta_1 + b_2 Z_i \text{ and } B_i = e^{b_0 + b_1 Z_i} \ ,
\end{aligned}
\tag{5}
$$

where k is the number of counties of the state and n_i is the number of census tracts (or towns) in county i.

The explanatory variable Mix is the proportion of white population in the census tract (town, respectively). We use the concentration of whites to measure the population mix because empirical evidence in Table 5 reveals that many Massachusetts counties have a very small nonwhite population. By including Mix, we are implicitly taking into account some location preferences that may be related to ethnic or network features. The idea, as described in Gabriel and Rosenthal (2004) and Zenou (2009), is to represent the greater or lesser willingness of people to interact with each other, even with some practice of discrimination. For instance, we can test to what extent individuals tend to cluster with other individuals belonging to the same ethnicity in order to share cultural habits.

As for the income variable Inc, we are alternately adopting per capita GDP when running the regression by exploiting the data from the town level and the median household income when focusing on the census tract data.[18]Income represents an interesting feature of the social peer effects and helps to capture the potential existence a specific economic territorial reality. In this respect, local per capita GDP is the most representative measure.

Another possible measure of peer effects is level of education $Educ$. Again we refer to the social network literature (for instance, Calvo-Armengol, Patacchini and Zenou, 2009). Individuals often place a high priority on the possibility of setting in areas where they share the same cultural interests with their neighbors. One way of successfully representing this dimension is to refer to the degree of education: persons with a higher degree of education may be more prone to settling in areas with others who have the same level of education in order to share common interests and, eventually, fuel joint cultural initiatives. In our econometric exercise, to take into consideration the composition of the education level of the population in a territorial unit, we create a synthetic indicator to take educational composition into account. Our indicator $Educ$ ranks the territorial units according to the education level of the population. We rank all units first according to the relative frequency of persons (older than 25 years in the year 2000) who have a high degree of education and second based on the relative frequency of persons who have an elementary degree of education. Then, for each territorial unit, we compute $Educ$ as the difference between the higher ranking minus the elementary ranking. In this way, this index allows us to

[18]Income data refer to 1999 as recorded in the 2000 Census. For census tracts we select the median household (which closely approximates per capita GDP) because the per capita GDP series is missing for census tracts.

take into account to what extent a unit may emerge as relatively highly educated with respect to other units in Massachusetts.

The covariate *Age* represents the ratio of the population between 18 and 64 years in the year 2000 (namely, working age population) over the total population of each territorial unit. In this respect, this indicator captures the relative concentration of potential workers in a territorial unit of reference.

We complete the analysis by also including in our equation the interaction terms between water area Z, income, education, age, and distance from Boston. These interaction terms can capture a degree of potential attractiveness[19] of each selected territorial unit embedding a sort of mass effect for a few select features (smoothed by the physical distance from Boston). They are expected to amplify the attractiveness of a destination privileging one of the environmental priorities that residents are searching for despite the physical distance from the CBD.

Concerning the spatial dimension, the Gamma model in (5) attempts to take into account the spatial dependence in each county by means of the state variables X_i. In our exercise, we are assuming that the land organizational structure of each county is independent of the other counties, but census tracts (or towns) belonging to the same county are characterized by very similar features. The idea is to highlight some common and institutional factors shared by census tracts (towns) inside the county, but these factors may differ across counties as, for instance, the land use regulation. It is also true that the townships belonging to the same county may share specific natural endowments that other townships in other counties do not share (e.g., Dukes County is an island). In other words, towns belonging to the same county share some common features that can be associated with fixed effects embedded in Z's and random effects represented by unobserved X's. Therefore, the population densities of different counties behave as independent random variables, whereas the spatial dependence between population densities in the county i is modeled via X_i.

We obtain several features of density of population Y_{ij} through the conditional expectation results and some standard properties of the Gamma distribution. For $\alpha > 1$ we have

$$\mathrm{E}(Y_{ij}) = \frac{\alpha}{\alpha-1}\mathrm{e}^{b_0+b_1 Z_i+b_2 Z_i D_{ij}+\beta_1 D_{ij}+\sum_{s=2}^{5}\beta_s V_{ij}^{(s)}+\sum_{s=3}^{5}\beta_{3+s}V_{ij}^{(s)}D_{ij}}, \tag{6}$$

whereas for $\alpha > 2$ we have the following:

$$\mathrm{Var}(Y_{ij}) = \frac{\alpha-1+\theta}{(\alpha-2)\theta}\times \mathrm{E}^2(Y_{ij}) \ , \tag{7}$$

$$\rho(Y_{ij}Y_{ih}) = \frac{\theta}{\theta+\alpha-1} \ , \tag{8}$$

[19]Enhancing one of the specific characteristics of the peer effect determinants.

since

$$
\begin{aligned}
\mathrm{E}\left(X_i^{-r}\right) &= \frac{\alpha^r B_i^r}{(\alpha-1)\times(\alpha-r)} \qquad \text{for } \alpha > r \text{ and } r = 1,2\ , \\
\mathrm{E}\left(Y_{ij}^r\right) &= E(E(Y_{ij}^r|X_i)) \\
&= \frac{\theta(\theta+1)\cdots(\theta+r-1)e^{r(\beta_1 D_{ij}+\sum_{s=2}^{5}\beta_s V_{ij}^{(s)}+\sum_{s=3}^{5}\beta_{3+s}V_{ij}^{(s)}D_{ij})}}{\theta^r}\,\mathrm{E}(X_i^{-r}) \qquad \text{for } r=1,2\ , \\
\mathrm{E}(Y_{ij}Y_{lh}) &= \begin{cases} \mathrm{E}(\mathrm{E}(Y_{ij}Y_{ih}|X_i)) = \mathrm{E}(\mathrm{E}(Y_{ij}|X_i)\,\mathrm{E}(Y_{ih}|X_i)) & \text{if } i=l \\ \mathrm{E}(Y_{ij})\,\mathrm{E}(Y_{lh}) & \text{if } i \neq l\ . \end{cases}
\end{aligned}
$$

We learn from (6) that the unconditional expectation of Y_{ij} describes a log-linear regression model that includes the local predictors D_{ij} and $V_{ij}^{(s)}$, the global predictor Z_i, and some interaction terms. The variance of Y_{ij} at (7) is quadratic in the mean, and the correlation $\rho(Y_{ij}Y_{ih})$ between the population densities of two cities in the same county is always positive and equal for all cities and counties. The shape parameters θ and α measure the intensity of the relationship between Y_{ij}, Y_{ih} within counties and introduce a tool for territorial differentiation. In particular, for large values of α, the state variables X_i's are concentrated around a fixed effect $B_i = b_0 + b_1 Z_i$, and the correlation of the Y_{ij} within counties approaches zero (see equation (8)). Similarly, small values of α represent a strong positive relationship of densities among cities in the same county, but a stronger heterogeneity among counties. For the α value to be equal, the larger the θ, the less concentrated the Y_{ij}'s (see (7)) and the larger the dependence between Y_{ij} and Y_{ih} (see (8)). When the county variable X_i concentrates around B_i, (i.e., only the water area Z_i has a discriminating impact on population distribution within each county), then other kinds of factors must be considered as potential discriminatory devices (here, represented by the parameters shaping the distribution function). Conversely, when those factors are practically identical across census tracts (respectively, towns) within the same county, it is less likely that the population density of one census tract (town) will be differentiated. In this second case, the individual preferences are less stringent.

Alternately, we can measure the dependence between the population densities of two census tracts (towns) in a county using Kendall's τ. In Appendix D we prove that for $\theta = 1$, $\tau(Y_{ij}, Y_{ih})$ is given by $\tau(Y_{ij}, Y_{ih}) = 2/(2+\alpha)$. For $\alpha \to \infty$, both ρ and τ approach zero as $1/\alpha$, but ρ approaches zero faster than τ does. Nevertheless, in Model (5) the dependence between Y_{ij}, Y_{ih} is not linear and thus could not be detected only by ρ. Finally, Kendall's τ allows for the detection of the dependence even if α is less than or equal to two, whereas some of the mean values of $Y_{ij}, Y_{i,h}$ involved in equations (6)-(8) do not exist for $\alpha \leq 2$.

4.2 Likelihood specifications of the Gamma-Gamma model

The next step is to define a suitable representation of the likelihood to estimate population density in accordance with the statistical framework we have developed. Let $(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{V}, \boldsymbol{Z})$ be the collection of the data $(Y_{ij}, D_{ij}, V_{ij}^{(s)}, Z_i)$ observed for every $j = 1, \ldots, n_i$ and $i = 1, \ldots, k$ and let $\boldsymbol{\beta} = (b_0, b_1, b_2, \beta_1, \ldots, \beta_8)$ be the vector of the 11 regression parameters.

The likelihood function $L(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{V}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha)$ of the parameters $(\boldsymbol{\beta}, \theta, \alpha)$ based on observed data $(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{V}, \boldsymbol{Z})$ can be obtained by integrating the conditional joint probability densities of $Y_{ij}|X_i$ over all the random state variables X_i. In addition, for $\theta \neq 1$, the final form of the likelihood $L(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha)$ is too complicated to work with.[20] Therefore, it is preferable to consider the following alternate model:

$$
\begin{aligned}
Y_{i,j}|w_i &\sim \text{Gamma}(\theta, \theta w_i/\mu_{ij}) \text{ independent for all } j \\
w_i &\sim \text{Gamma}(\alpha, \alpha) \text{ independent for all } i \\
\mu_{ij} &= \mathrm{e}^{b_0 + b_1 Z_i + b_2 Z_i D_{ij} + \beta_1 D_{ij} + \sum_{s=2}^{5} \beta_s V_{ij}^{(s)} + \sum_{s=3}^{5} \beta_{3+s} V_{ij}^{(s)} D_{ij}},
\end{aligned}
\tag{9}
$$

which gives rise to the same likelihood $L(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{V}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha)$. Considering that the random factors X_i and w_i are unobservable, the two models are not distinguishable, and every estimate of parameters $(\boldsymbol{\beta}, \theta, \alpha)$ obtained using the likelihood function should be the same. Owing to its increased simplicity, we use Model (9) to develop an estimation procedure for $(\boldsymbol{\beta}, \theta, \alpha)$.[21]

In our econometric exercise, we estimate four alternative versions of Model (9). The first, labeled *Model 0*, is a reduced form of (9) where the unique covariate of the regression is the Euclidian distance D_{ij} (i.e., we let $b_1 = b_2 = \beta_2 = \cdots = \beta_8 = 0$ in (9)). We use *Model 0* for two goals: first to identify the CBD (see Table 2 and the discussion in the first paragraph of Section 4) and second to benchmark the goodness of fit of the augmented model with the other covariates. The other three models employ different indicators of environmental preferences of individuals as explanatory variables of population density: our first selected model (henceforth referred to as *Model I*) focuses on ethnic, age and education indicators and is obtained by plugging $\beta_3 = \beta_6 = 0$ into Model (9). The second model (henceforth referred to as *Model II*) addresses the ethnic, age and income indicators and is obtained by letting $\beta_4 = \beta_7 = 0$ in (9). The third model (named *Model III*) merges the salient features of the two previous Models *I* and *II* by focusing on all of the environmental indicators previously selected and coincides with Model (9). Thus, our idea is to test the statistical robustness of the results by selecting different specifications that either jointly consider or isolate a few of the various indicators and therefore center on specific determinants.

4.3 Prior specifications and Bayesian estimation

We use a Bayesian approach to estimate $\boldsymbol{\beta}, \alpha, \theta$.

From a Bayesian perspective, the unknown parameters are understood as random variables with a *prior* joint distribution, say, π, and the statistical problem consists of updating π by computing a *posterior* joint conditional probability of $\boldsymbol{\beta}, \alpha, \theta$, given the data $\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{V}, \boldsymbol{Z}$. The posterior joint distribution is then summarized in a simple way, typically with posterior means giving rise to a point estimate of the unknown parameters. Moreover, the associated posterior standard errors and a $\gamma 100\%$ credible interval[22] for the unknown parameters are computed. Usually, both the joint and the marginal posterior distributions of all the unknown parameters do not have a closed form. Hence, one may use some Markov Chain Monte

[20] A representation for $L(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha)$ is given by $L(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{V}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha) = \prod_{i,j} y_{ij}^{\theta-1} \times \dfrac{\theta^{n\theta} \alpha^{k\alpha} \prod_{i=1}^{k} \Gamma(n_i \theta + \alpha)}{\Gamma(\theta)^n \Gamma(\alpha)^k} \times \prod_{i=1}^{k} B_i^{\alpha} \Big(\alpha B_i + \theta \sum_j y_{i,j} \mathrm{e}^{a_i D_{ij} + \sum_{s=2}^{5} \beta_s V_{ij}^{(s)} + \sum_{s=3}^{5} \beta_{3+s} V_{ij}^{(s)} D_{ij}} \Big)^{-(n\theta+\alpha)}$,

[21] The parameter restriction shape = scale = α for the Gamma distribution of the frailty terms w_i (which results in $\mathrm{E}(w_i) = 1$) ensures that Model (9) is identifiable.

[22] In Bayesian statistics, a $\gamma 100\%$ credible interval for a parameter η is given by $q_{(1-\gamma)/2} < \eta < q_{(1+\gamma)/2}$, where $q_{(1-\gamma)/2}, q_{(1+\gamma)/2}$ are posterior quantiles of η.

Carlo algorithms to simulate and summarize them.

Let $\boldsymbol{w} = (w_1, \ldots, w_k)$ be the vector of the unobservable "frailties", $(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{V}, \boldsymbol{Z})$ the set of the "complete data" and $L(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{V}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha)$ the "complete" likelihood given by

$$L(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{V}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha) = \frac{\theta^{n\theta}\alpha^{k\alpha}}{\Gamma(\theta)^n\Gamma(\alpha)^k} \prod_{i,j}\left(\mu_{ij}^{\theta} y_{ij}^{\theta-1}\right) \mathrm{e}^{-\sum_{i=1}^{k} w_i(\theta\sum_j \mu_{ij} y_{ij}+\alpha)} \prod_i w_i^{n_i\theta+\alpha-1} \tag{10}$$

with $n = \sum_{i=1}^{k} n_i$. From now on, we will work with this complete likelihood (10) and handle unknown frailties $\boldsymbol{w}$ as unknown parameters to estimate.

With regard to the prior, we selected "noninformative" priors for $\boldsymbol{\beta}, \alpha, \theta$ to represent our vague prior knowledge of these factors. A priori $\boldsymbol{\beta}, \alpha$ and θ are assumed to be independent. In particular, the regression parameters in $\boldsymbol{\beta}$ are a priori independent normal random variables with large variances: $b_r, \beta_r \sim \mathrm{N}(0, 10{,}000)$, and both the shape parameter α and the rate θ are (independent) exponential with a rate 0.2: this choice for α, θ implies a large prior variance for α, θ.[23]

4.4 Results for the Massachusetts case study

To deal with tractable values, we standardized the regressors, excluding the constant term, by subtracting the sample mean and dividing by the sample standard deviation. All the inferences were coded in the JAGS (Just Another Gibbs Sampler) software package by Plummer (2010), which is designed to work closely with the R package in which all statistical computations and graphics were performed. On the whole, $750,000$ iterations for three chains were run for the unknown parameters $\boldsymbol{\beta}, \alpha, \theta$ and frailties $w_1, \ldots, w_{14}$, and the first $250,000$ were discarded as burn-in. One out of every 100 values simulated after the burn-in was kept for posterior analysis, for a total of $5,000$ simulations saved per chain. We selected one final sample among these three. The convergence diagnostics such as those available in the R package CODA were computed (Gelman, Geweke, Heidelberger and Welch stationarity test, interval halfwidth test) for all parameters, indicating that convergence may have been achieved.

4.4.1 The census tract case

In this subsections we consider the data for all of the $1,361$ census tracts in Massachusetts (from the 2000 Census) and report the results of fitting Models *I, II and III* to the data.

Table 4(a) shows the posterior mean, standard deviation (first row) and 95 percent credible intervals (second row) for the effective parameters $\boldsymbol{\beta}, \alpha, \theta$. The third row of Table 4(a) shows the posterior predictive p-values of regression parameters: $p_{b_i} = \min\{P(b_i > 0|\boldsymbol{Data}), P(b_i < 0|\boldsymbol{Data})\}$, $p_{\beta_i} = \min\{P(\beta_i > 0|\boldsymbol{Data}), P(\beta_i < 0|\boldsymbol{Data})\}$. We can draw conclusions about the association between each predictors and Y if the p-value is low (for example, less than 5 percent.) Conversely, p_{β_i} (p_{b_i}) is close to 0.5 when β_i (b_i) is concentrated around zero.

Figure 1(a) displays 95 percent posterior credible intervals of the gamma county frailties $\boldsymbol{w}$, with the

[23] The most commonly used noninformative prior law for a rate parameter is Gamma(ν, ν) with a small ν. However, our exponential prior choice allows for a better mixing of the MCMC algorithm. Moreover, in a sensitivity analysis, we compared the results of fitting the Gamma-Gamma models using both of these noninformative priors and obtained very similar results for the estimates of the regression parameters.

(a) Census tract data

	Model I	*Model II*	*Model III*
b_0 (Intercept)	1.339 (0.173) (0.991, 1.676) 0	1.143 (0.168) (0.796, 1.457) 0	1.199 (0.166) (0.852, 1.307) 0
b_1 (Z_i)	-0.081 (0.020) ($-0.119, -0.041$) 0	-0.074 (0.019) ($-0.111, -0.035$) 0	-0.080 (0.019) ($-0.118, -0.067$) 0
b_2 ($Z_i \times Dist$)	-0.016 (0.021) ($-0.058, 0.025$) 0.217	-0.009 (0.020) ($-0.049, 0.005$) 0.330	-0.021 (0.020) ($-0.059, -0.008$) 0.154
β_1 ($Dist$)	-0.964 (0.101) ($-1.163, -0.774$) 0	-1.012 (0.087) ($-1.187, -0.847$) 0	-1.051 (0.092) ($-1.233, -0.875$) 0
β_2 (Mix)	-0.307 (0.035) ($-0.378, -0.239$) 0	-0.305 (0.032) ($-0.368, -0.243$) 0	-0.240 (0.033) ($-0.306, -0.176$) 0
β_3 (Inc)	–	-0.615 (0.030) ($-0.671, -0.556$) 0	-0.437 (0.042) ($-0.518, -0.355$) 0
β_4 ($Educ$)	-0.544 (0.033) ($-0.611, -0.479$) 0	–	-0.262 (0.043) ($-0.347, -0.234$) 0
β_5 (Age)	0.441 (0.023) (0.397, 0.488) 0	0.319 (0.023) (0.276, 0.364) 0	0.370 (0.024) (0.323, 0.417) 0
β_6 ($Inc \times Dist$)	–	-0.314 (0.028) ($-0.368, -0.261$) 0	-0.258 (0.041) ($-0.339, -0.176$) 0
β_7 ($Educ \times Dist$)	-0.183 (0.026) ($-0.233, -0.132$) 0	–	-0.044 (0.037)9 ($-0.117, -0.01$) 0.119
β_8 ($Age \times Dist$)	-0.013 (0.020) ($-0.053, 0.027$) 0.249	0.008 (0.018) ($-0.026, 0.044$) 0.337	0.003 (0.018) ($-0.032, 0.040$) 0.420
$\alpha_{frailty}$	3.059 (1.362) (1.173, 6.467)	3.986 (1.842) (1.437, 8.569)	3.737 (1.751) (1.385, 8.108)
θ	1.546 (0.055) (1.442, 1.658)	1.624 (0.059) (1.511, 1.741)	1.663 (0.059) (1.548, 1.783)
DIC (pD)	6250.3 (30.5)	6169.3 (29.8)	**6136.1***** (33.8)
% outliers	4.34%% (9.18%)	4.26% (9.26%)	4.63% (9.33%)

(b) Town data

	Model I	*Model II*	*Model III*
b_0 (Intercept)	-0.614 (0.213) ($-1.074, -0.238$) 0.0018	-0.606 (0.202) ($-1.047, -0.235$) 0.001	-0.618 (0.212) ($-1.081, -0.231$) 0.0012
b_1 (Z_i)	-0.042 (0.132) ($-0.325, 0.204$) 0.3848	-0.052 (0.130) ($-0.329, 0.183$) 0.3504	-0.046 (0.133) ($-0.338, 0.187$) 0.3706
b_2 ($Z_i \times Dist$)	0.124 (0.142) ($-0.126, 0.429$) 0.1886	0.102 (0.139) ($-0.140, 0.407$) 0.2312	0.125 (0.142) ($-0.125, 0.429$) 0.1886
β_1 ($Dist$)	-1.420 (0.169) ($-1.759, -1.099$) 0	-1.398 (0.166) ($-1.726, -1.087$) 0	-1.426 (0.170) ($-1.762, -1.107$) 0
β_2 (Mix)	-0.581 (0.064) ($-0.710, -0.456$) 0	-0.596 (0.064) ($-0.725, -0.478$) 0	-0.574 (0.065) ($-0.705, -0.451$) 0
β_3 (Inc)	–	-0.177 (0.053) ($-0.277, -0.070$) 0.0002	-0.077 (0.076) ($-0.224, 0.075$) 0.1552
β_4 ($Educ$)	-0.181 (0.050) ($-0.281, -0.086$) 0.0002	–	-0.133(0.070) ($-0.269, 0.005$) 0.0302
β_5 (Age)	-0.197 (0.045) ($-0.282, -0.109$) 0	-0.226 (0.046) ($-0.315, -0.135$) 0	-0.207 (0.046) ($-0.297, -0.114$) 0
β_6 ($Inc \times Dist$)	–	-0.027 (0.042) ($-0.109, 0.057$) 0.2594	-0.032 (0.069) ($-0.168, 0.102$) 0.3176
β_7 ($Educ \times Dist$)	-0.005 (0.046) ($-0.093, 0.084$) 0.463	–	0.014 (0.072) ($-0.131, 0.155$) 0.4258
β_8 ($Age \times Dist$)	-0.194 (0.041) ($-0.275, -0.115$) 0	-0.208 (0.040) ($-0.289, -0.130$) 0	-0.195 (0.041) ($-0.275, -0.117$) 0
$\alpha_{frailty}$	2.824 (1.601) (0.927, 6.963)	3.070 (1.718) (0.972, 7.551)	2.858 (1.575) (0.923, 6.763)
θ	1.953 (0.140) (1.688, 2.239)	1.932 (0.138) (1.673, 2.214)	1.945 (0.142) (1.679, 2.224)
DIC (pD)	**264.1***** (33.1)	266.6 (32.3)	267.6 (35.8)
% outliers	3.70%% (9.40%)	3.42% (9.12%)	3.70% (9.69%)

Table 4: Summaries of the posterior distributions of $\boldsymbol{\beta}, \alpha, \theta$, DIC values (with pD) and percent of units (% outliers) which have posterior p-value less than 2.5 percent (5 percent), with census tract and town data, from different Models. Posterior means of $\boldsymbol{\beta}, \alpha, \theta$ are followed by (standard deviations) in the first row, 95 percent credible intervals are shown in the second row, and the posterior predictive p-value of the regression parameters are shown in the third row.

value of their posterior mean as the $\times$ symbol and the median as a solid point. To assess the goodness of fit and compare Models *I, II, III*, we use the DIC (see the bottom of Table 4(a)) and the *posterior predictive model-checking* approach, both illustrated, for example, in Ntzoufras (2009, Chap. 10).

Illustrating the posterior predictive model-checking in our context, let $f(y^*_{ij}|\boldsymbol{Data})$ be the marginal "posterior predictive" distribution of the population density of territorial unit j in county i, in Massachusetts:

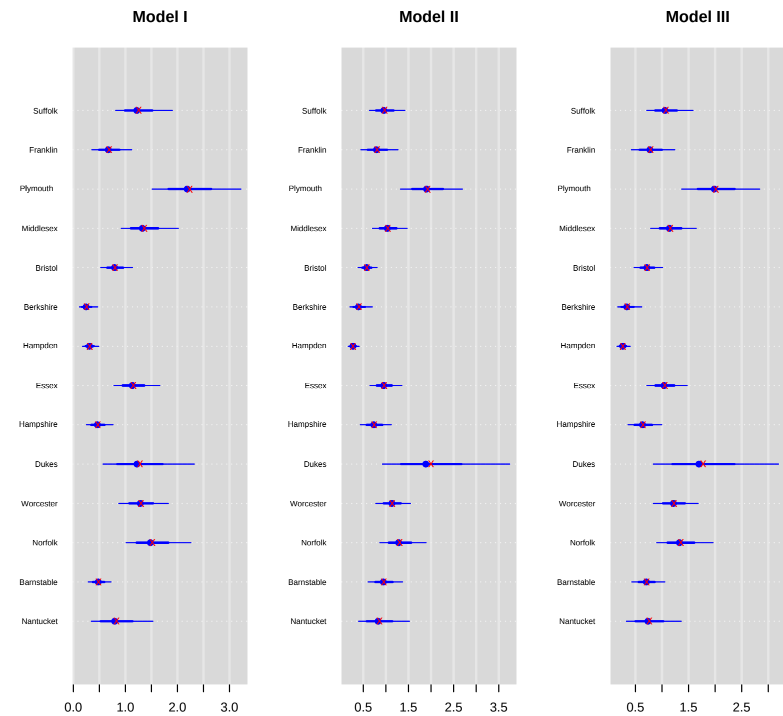
$$f(y^*_{ij}|\boldsymbol{Data}) = \int f(y^*_{ij}|\boldsymbol{\eta})\pi(\boldsymbol{\eta}|\boldsymbol{Data})\, d\boldsymbol{\eta}.$$

where, for simplicity, $\boldsymbol{\eta}$ denotes all the parameters: $\boldsymbol{\eta} = (\boldsymbol{\beta}, \theta, \alpha, \boldsymbol{w})$. We have simulated L draws $\boldsymbol{y}^*_{ij} = (y^*_{ij;1}, \ldots, y^*_{ij;L})$ from $f(y^*_{ij}|\boldsymbol{Data})$ for each unit in Massachusetts, and we compared the actual densities of the population y_{ij} in $\boldsymbol{Data}$ with the predicted densities of the population $\boldsymbol{y}^*_{ij}$. More specifically, we summarized each marginal posterior predictive distribution $f(y^*_{ij}|\boldsymbol{Data})$ by its 95 percent credible interval and defined the actual values of y_{ij}'s that lie outside of the 95 percent credible interval as possible outliers. Indeed, if an observed y_{ij} is in the tail of the predictive distribution, then it has a very small *posterior predictive p-value*[24] (less than 2.5 percent or less than 5 percent) which indicates a failure of the model for it. We then showed the posterior percentage ("% outliers") of the units for which *Models I, II, III* fail according to the p-value criterion in the last row of Table 4(a) as an overall measure of fit. We also graphed the posterior 95 percent credible interval of $f(y^*_{ij}|\boldsymbol{Data})$ as line plots in Figure 1(b) and placed the observed actual densities y_{ij} as solid dots. Finally, Figure 2(a) displays the scatterplot of *i)* the actual log population density y_{ij} (with circles), *ii)* the posterior expected census tract log densities for the reduced *Model 0* labeled by down-pointing triangles and *iii)* the posterior expected census tract log densities under (9) (with up-pointing triangles) versus the distance from Boston. In comparing the performance of the three previous models according to the DIC, at the census tract level, *Model III* turns out to be the most efficient one. However, the three models perform well and similarly in identifying the outliers. In this respect "% outliers" is equal for all of the models: around 4 percent at a 2.5 percent significance level and around 9 percent at 5 percent. Analyzing the outliers separately in each county, we found that our Gamma-Gamma models fail in the most remote and border counties: Franklin, Berkshire and Hampshire, where the proportion of outliers ranges between 13 percent and 31 percent of the total number of each county census tracts.

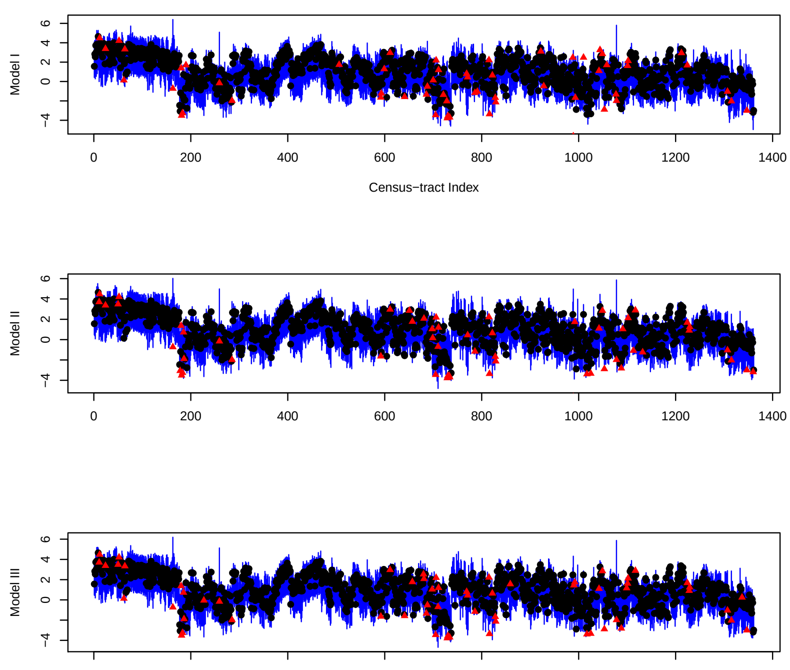
As for the regression coefficients, the three models provide consistent estimations. As we can deduce from Table 4(a), the significant terms in the prediction of density population Y are the Euclidean distance from Boston, the water area Z, and, individually, ethnic composition (Mix), income (Inc), education ($Educ$) and age (Age). The distance from Boston has a negative impact on the population density distribution (95% credible interval of $\beta_1 = (-1.233, -0.835)$). Additionally, the presence of large water areas is a clear deterrent for population concentrations, as expected because of the physical obstacles to construction.

Individuals also seem to be sensitive to ethnic composition. Boston turns out to be significantly

[24] The posterior predictive p-value is defined as the probability that an observation Y^*_{ij} governed by the posterior predictive distribution $f(y^*_{ij}|\boldsymbol{Data})$ is "*at least as extreme as*" the observed density population y_{ij}: $p^*_{ij} = \min\left\{P(Y^*_{ij} < y_{ij}), P(Y^*_{ij} > y_{ij})\right\}$.



(a) Counties gamma frailties $\boldsymbol{w}$



(b) log census tracts population log densities

Figure 1: In Figure (a): 95 percent posterior credible intervals of gamma county frailties $\boldsymbol{w}$ with census tract data. In Figure (b): posterior predictive 95 percent credible intervals of the census tracts log population density with actual census tracts log densities $\ln y_{ij}$ denoted by (black) solid dots. Suspect outliers denoted by (red) up-pointing triangles.

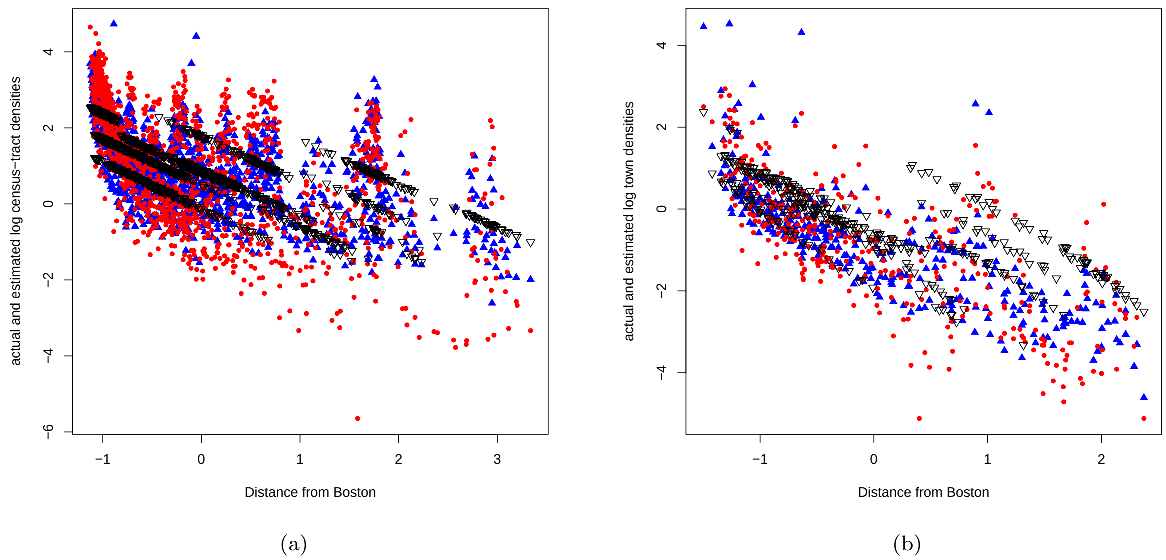


Figure 2: Scatterplot with census tract data in (a) and town data in (b). (1) Census tract real log densities denoted by (red) circles, (2) posterior expected census tract log densities for the reduced *Model 0* labeled with (black) down-pointing triangles and (3) posterior expected census tract log densities under *Model III* denoted with (blue) up-pointing triangles, versus distance from Boston.

more attractive for the nonwhite than for the white population and, therefore, nonwhite individuals display stronger preferences for settling in or around Boston. This issue is generally common to other US metropolitan areas (McMillen 2004) and may reflect the combination of two different forces. On the one hand, members of the nonwhite population (often immigrants or racially segregated citizens) usually display a stronger tendency for searching out and establishing strong social ties, as well as seeking out other individuals with the same ethnicity. In the CBD, nonwhite individuals are more likely to join (or interact with) other ethnic minorities; that is, they become a part of an ethnic social network. On the other hand, the lack of a complete assimilation process between different ethnicities, associated with the effect of strong income disparities may shape nonwhite individual preferences for settling in the proximity of the metropolitan area. To this extent, the income and education indicators (as well as the interaction terms between distance and income) reinforce the previous finding: the higher one's education and income level, the more willing one is to settle far from Boston. In this respect, the cohesion force of similar individuals is also captured by the age indicator. This indicator shows that a concentrated mass of working-age population is relatively good at attracting people who share similar characteristics. Still, we notice in Table 4(a) that, irrespective of the models we fit, the interaction terms between the distance from Boston and the size of the water area Z and Age have regression coefficients concentrated closely around zero.

In terms of coefficient sizes, distance is revealed to be the relatively the most important factor, since the magnitude of its coefficient corresponds to the sum of all of the other covariates. The scatterplots of real and estimated log densities in Figure 2(a), however, infer that the distance factor alone (black down-pointing triangles in the graphs) is not able to capture the highest and lowest true densities (red circle points). The distance factor in *Model 0* is able to picture the trend but not the variance, and because of the spread of the population density distribution, we need to address this variance. Figure 2 shows that our *Model III* (up-pointing blue triangles), which includes the proxies for subjective preferences, performs well at capturing the highest densities.

The analysis of the statistical significance of gamma frailties $\boldsymbol{w}$ confirms the importance of county random effects. First, the posterior means estimates of α corresponding to the different models in Table 4(a) show a strong posterior evidence of a high degree of heterogeneity among the counties and a posterior positive dependence between two census tracts belonging to the same county: for all *Models I, II, III* the posterior mean of α given the data is smaller than prior mean $\mathrm{E}(\alpha) = 1/0.2 = 5$. Moreover, Figure 1(a) depicts that the $\boldsymbol{w}$ values are almost all concentrated far away from one. The counties of Suffolk and Essex are the exception for *Models II and III*, in which the income covariate appears. These counties are close and share borders, hence the low degree of heterogeneity may be interpreted as a signal that they would be best evaluated as a unique territorial unit.

On the whole, our results reinforce the validity of our choice to introduce the concept of subjective distance rather than simple Euclidean distance, and the Bayesian technique is revealed to be suitable for our purpose.

4.4.2 The case of a selected sample of towns in Massachusetts

In the second econometric exercise, we focus our analysis on the sample of 351 towns in Massachusetts (according to data released by the Census Bureau). We perform the exercise by running *Models 0, I, II,* and *III* as in the case of census tract data.

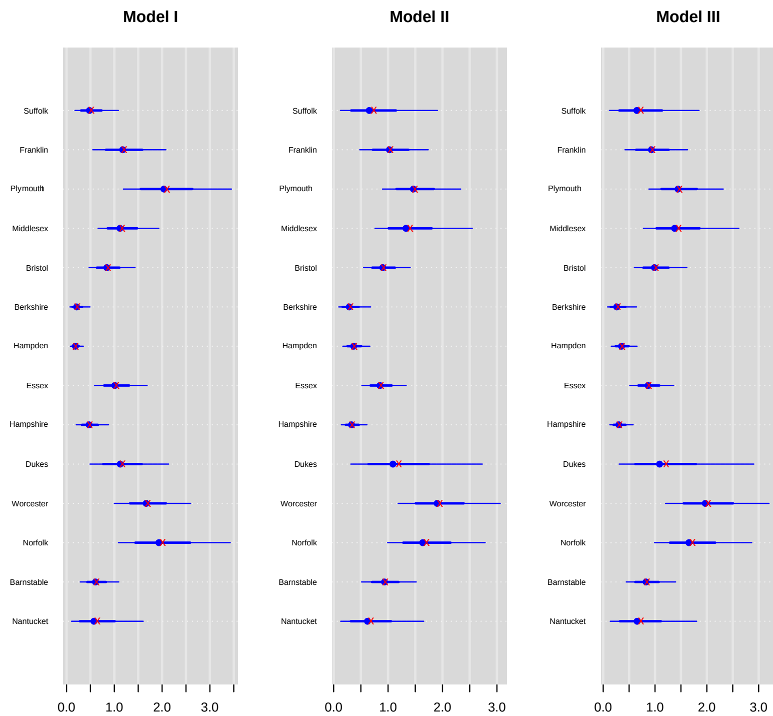
The three specifications convey the same results as shown in Table 4(b): the size of the estimated coefficients is almost identical across the three specifications as is the statistical significance of the selected covariates. However, according to the DIC criterion, *Model I* (including distance, ethnic composition, education and age as local covariates) is preferred. Once more, the distance from Boston is negatively associated with density size, and the magnitude of the correspondent coefficient suggests that this effect is the largest among the group of covariates we consider. Looking at *Models II* and *III* we find, again, that the ethnic, education and income covariates display negative and significant coefficients confirming that high-earning, educated and white persons display preferences to settle far from Boston. Despite the results of the census tract data, in the estimations with town data the water area covariate is not significant at all, and the age composition is shown to have a negative and statistically significant coefficient. These results reinforce the portrait of the selected group aiming at living far from Boston as composed mostly of working-age individuals (rather than retired individuals) who are likely to earn a high income. We consider the absence of statistical significance of the water areas to be due mostly to the selection criteria of the sample. We are exclusively focusing on urban territories that may be located in more or less dense urban areas in Massachusetts. Instead, this covariate is more suitable for capturing the structural heterogeneity that exists in the composition of the territorial rural land; water areas are likely to represent a latent (discriminating) feature to be considered when analyzing census tract data covering the entire territory of Massachusetts. The scatterplot of the real and estimated densities in Figure 2(b) for *Model I* confirms that our model (up-pointing blue triangles) provides a better fit of the true density distribution in the upper and lower tails of the distribution. In contrast, *Model 0*, which accounts only for for the physical distance from Boston, mostly captures the trend but without precisely fitting the observations in the tails. In addition, when comparing the efficiency of the estimations of our models with the two datasets, our estimations permit a better fit of the low-density observation (in the right tails) using the town data, in contrast to using the census tract data. In our opinion, this outcome is due mostly to the lower degree of heterogeneity with town data, which drives the covariates to better match the true data. To complete the information associated with our estimates, Figures 3(a) and (b) include the graphs referring to the 95 percent credible intervals of the gamma frailties $\boldsymbol{w}$ and of the marginal posterior predictive distributions $f(y^*_{ij}|\textbf{town Data})$ under *Models I, II, III*, respectively. Again, the fit of our models is good because the statistics in Table 4(b) replicate the pattern described for the census tract data and the outliers (at 2.5 percent significance level) make up less than 3.7 percent of the sample for each one of *Models I, II, III.*

5 Concluding remarks

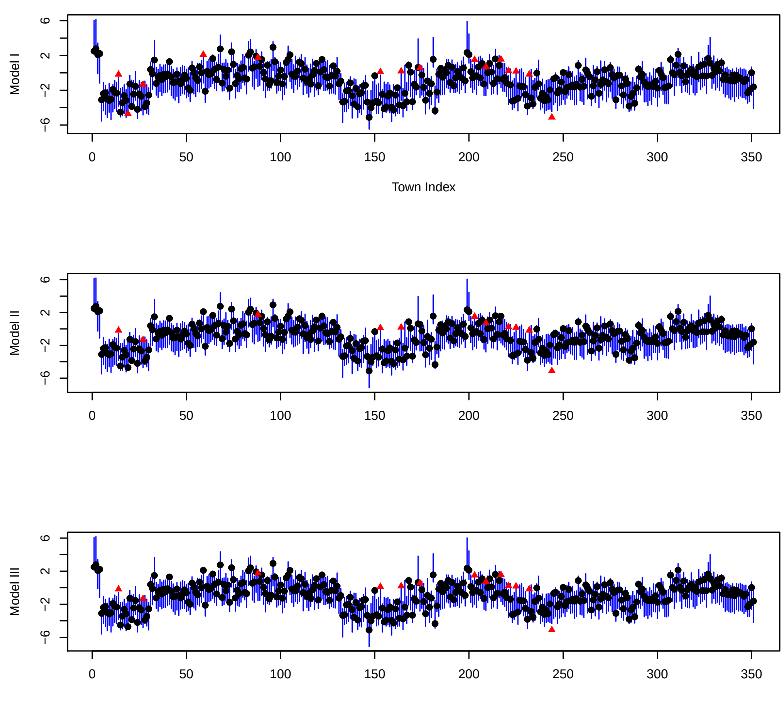
In our paper, we propose a probabilistic approach to estimating the population density distribution for a regional territory characterized by the coexistence of rural and urban land plots. We develop a general framework by working with a probabilistic distribution built around the association of individual preferences for different location options. The population density distribution and the distance function are modeled as random variables. We adopt Kendall's τ to enhance the importance of the individual preferences to settle close to the center. The empirical strategy we follow is pegged on the statistical properties of the Gamma function, which allow us to take into account the heterogeneity of the space as claimed in spatial theory. We then perform an empirical exercise to test the fit of our strategy of analysis by exploiting the information from census tract and town data in Massachusetts. One novelty of this approach is the modeling of a subjective measure of distance that merges the physical distance between two locations with certain measures of individual preferences. According to the data at hand, in our exercise, income, age, education and ethnic composition are jointly as important as the physical distance from Boston for fitting the population density distribution.

In future developments of this study, we envision extending this empirical exercise to other case studies, as well as proposing further developments of the Massachusetts case by prioritzing the historical dimension. The second idea would be to test the potential dynamic evolution of the goodness-of-fit of our framework in the context of shifting the land markets and household structure. The potential deviation, if any, from the current specification would provide some suggestions as to the interpretation of variation of habits or preferences among individuals across time.

Another natural extension is to think of a polycentric spatial configuration. In accordance with the current framework, the extension to a polycentric setting would be inspired by the Löschian central place theory to generate a proposal of urban hierarchy that overcomes the limits of a monocentric distribution. In the spirit of Lösch, the ranking of cities relies on a process of maximization of consumer welfare, whereby the need to travel for any good is minimized. In this sense, we have cities that provide the highest-order services (or goods) and have a huge hinterland (namely, first-order places) followed by other cities (namely, second-order and third-order cities) supplying progressively fewer services. The individual preferences in terms of proximity to one or more services drive the individual location decisions



(a) Gamma county frailties $\boldsymbol{w}$



(b) Town population log densities

Figure 3: Figure (a): 95 percent posterior credible intervals of gamma county frailties $\boldsymbol{w}$, with town data. Figure (b): posterior predictive 95 percent credible intervals of the town population log densities with actual town log densities $\ln y_{ij}$ denoted by (black) solid dots. Suspect outliers denoted by (red)up-pointing triangles.

(e.g., to live in a first-order or second-order city). The mobility of individuals across different locations is due to their desire to access the services they demand. This movement creates potential links across the different order of towns and shapes the land structure. As a consequence of the presence of these spatial interactions, our framework should be modified by including some spatial filtering techniques such as those presented in a Bayesian framework by Crespo-Cuaresma and Feldkircher (2010).

Furthermore, the empirical results for Massachusetts suggest the need for modeling a spatial autocorrelation component between counties. In fact, the population densities of close counties may behave in a similar way. Hence, further research should more deeply characterize the basic components of the spatial structure of the territorial units.

Another important issue we did not address here is the stability over time of a Gamma-Gamma model to measure the population density in terms of the subjective distance function. The population mix can change gradually over time, or occasionally experience rapid changes, such as a sizeable immigration event or a rapid expansion of access to the Internet, which may have a dramatic impact on an individual's perceived metric distance. As a consequence, we aim to tackle this problem from a dynamic viewpoint.

Acknowledgments

We thank the editor, S. Brakman, and three anonymous referees for providing useful suggestions. We are also grateful to J. Caballé, G. Hewings, S. MacEachern, C. Nicodemo, J. Parr, S. Rosenthal and the participants at the Third Spatial Econometric World Conference (Barcelona), XXXIV SAEe (Valencia), 58th North American Regional Science Council meeting (Denver) as well as participants at the seminar at Université de Luxembourg and CUNY for comments. Special thanks go to M.A. Garcia-Lopez for professional assistance with the ArcView software. The usual disclaimer applies. R. Nicolini's research was supported by a *Ramón y Cajal* contract. Financial support from research grants 2009SGR600 and SEJ2008-01850 is acknowledged.

References

Alperovich, Gershon. 1982. 'Density gradients and the identification of the central business district,' *Journal of Urban Economics*, 17, 313–320.

Anas, Alex, Arnott, Richard, and Small, Kenneth A. 1998. 'Urban spatial structure,' *Journal of Economic Literature*, 36, 1426–1464.

Anselin, Luc 2010. 'Thirty years of spatial econometrics,' *Papers in Regional Science*, 89, 3–25.

Beckmann, Martin 1971. 'Von-Thünen Revisited: A Neoclassical Land Use Model,' *The Swedish Journal of Economics*, 74, 1–7.

Bracken, Ian and Martin, Dave 1988. 'The generation of spatial population distribution from census centered data,' *Environment and Planning A*, 21, 537–543.

Calvo-Armengol, Antoni, Patacchini, Eleonora, and Zenou, Yves 2009. 'Peer Effect and Social Networks in Education,'*Review of Economic Studies*, 764, 1239–1267.

Clark, Colin 1951. 'Urban Population density,' *Journal of the Royal Statistical Society*, 114, 490–494.

Crespo Cuaresma, Jesus, and Feldkircher, Martin 2012. 'Spatial Filtering Model Uncertainty and the speed of Income Convergence in Europe,' *Journal of Applied Econometrics*, doi: 10.1002/jae.2277.

Dana, Rose-Anne, and Scarsini, Marco 2007. 'Optimal risk sharing with background risk,' *Journal of Economic Theory*, 311, 152–176.

Firth, David 1988. 'Multiplicative errors: Log-normal or gamma?,' *Journal of the Royal Statistical Society: Series B*, 50, 266–268.

Frankena, Mark 1978. 'A bias in Estimating Urban Population Density Functions,' *Journal of Urban Economics*, 5, 35–45.

Fujita, Masahisa and Thisse, Jacques-François 2002. *Economics of Agglomeration. Cities, Industrial Location and Regional Growth.* Cambridge: Cambridge University Press.

Gabriel, Stuart and Rosenthal, Stuart 2004. 'Quality of the Business Environment versus quality of life: do firms and households like the same cities?,' *Review of Economics and Statistics*, 86, 438–444.

Glaeser, Edward 2008. *Cities, Agglomeration and Spatial Equilibrium.* Oxford: Oxford University Press.

Helsley, Robert and Strange, William. 2007. 'Urban interactions and spatial structure,' *Journal of Economic Geography*, 7, 119–138.

Helsley, Robert and Zenou, Yves 2011. *'Social Networks and Interactions in Cities,'* CEPR Discussion Paper No. 8260.

Law, Jane and Haining, Robert 2004 'A Bayesian Approach to Modeling Binary data: The Case of High-Intensity Crime Areas,' *Geographical Analysis*, 36, 197–216.

Lehmann, Erik Leo 1966. 'Some concepts of dependence,' *The Annals of Mathematical Statistics*, 37, 1137–1153.

LeSage, James 1997. 'Bayesian Estimation of Spatial Autoregressive Models,' *International Regional Science Review*, 20, 113–129.

LeSage, James, and Pace, Kelley R. 2009. *Introduction to Spatial Econometrics.* Boca Raton, FL: CRC Press.

Mas-Colell, Andreu, Whinston, Michael D., and Green, Jerry R. 1995. *Microeconomic Theory.* New York: Oxford University Press.

McCullagh, Peter and Nelder, John A. 1989. *Generalized Linear Models.* London: Chapman and Hall.

McDonald, John F. 1987. 'The identification of Urban Employment Subcenters' *Journal of Urban Economics*, 21, 242–258.

McDonald, John F. 1989. 'Econometric studies of Urban population density: a survay,' *Journal of Urban Economics*, 26, 361–385.

McDonald, John F. and Bowman, H. Woods 1976. 'Some tests of alternative urban population density functions,' *Journal of Urban Economics*, 3, 241–252.

McMillen, Daniel P. 2004. 'Employment Densities, Spatial Autocorrelation, and Subcenters in Large Metropolitan Areas,' *Journal of Regional Science*, 44, 225–244.

Milgrom, Paul R. and Weber, Robert J. 1982. 'A Theory of auctions and competitive bidding,' *Econometrica*, 50, 1089–1122.

Mills, Edwin S. 1972. *Studies in the Structure of the Urban Economy*. Baltimore: Johns Hopkins Press.

Mills, Edwin S. and Peng Tan, Jee 1980. 'A comparison of urban population density functions in developed and developing countries,' *Urban Studies*, 17, 313–321.

Muth, Richard 1969. *City and Housing: the Spatial Pattern of Urban Residential Land Use*. Chicago: University of Chicago Press.

Nairn, Alasdair G.M. and O'Neill, Gerard J. 1988. 'Population Density Functions: A differential Equation Approach,' *Journal of Regional Science*, 28, 89–102.

Nelsen, Roger N. 1999. *An Introduction to Copulas*. Lecture Notes in Statistics, 139. New York: Springer-Verlag.

Ntzoufras, Ioannis 2009. *Bayesian Modeling Using WinBUGS*. Hoboken: Wiley-Blackwell.

Parr, John 1985a. 'A Population-Density Approach to Regional Spatial Structure,' *Urban Studies*, 22, 289–303.

Parr, John 1985b. 'The Form of the Regional Density Function,' *Regional Studies*, 19, 535–546.

Parr, John 2007: 'Spatial Definitions of the City: Four Perspectives', *Urban Studies*, 442, 381–392.

Parr, John, O'Neill, Gerard J., and Nairn, Alasdair G.M. 1988: 'Metropolitan Density Functions. A Further Exploration', *Regional Science and Urban Economics*, 18, 463–478

Parr, John and O'Neill Gerard 1989: 'Aspects of the lognormal function in the analysis of regional population distribution', *Environment and Planning A*, 21, 961–973.

Plummer, Martyn 2010. *JAGS Version 3.1.0: Just Another Gibbs Sampler*, URL http://mcmc-jags.sourceforge.net/

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing,* Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rappaport, Jordan 2009. 'The increasing importance of quality of life,' *Journal of Economic Geography*, 9, 779–804.

Song, Shunfeng H. 1996. 'Some Tests of Alternative Accessibility: A Population Density Approach,' *Land Economics*, 72, 474–482.

Wood, Gavin A. and Parr, John. 2005. 'Transaction Costs, Agglomeration Economics, and Industrial Location,' *Growth and Change*, 36, 1–15.

Zenou, Yves 2009. *Urban Labor Economics*. New York: Cambridge University Press.

A Empirical evidence on density population in Massachusetts; graphs and descriptive statistics

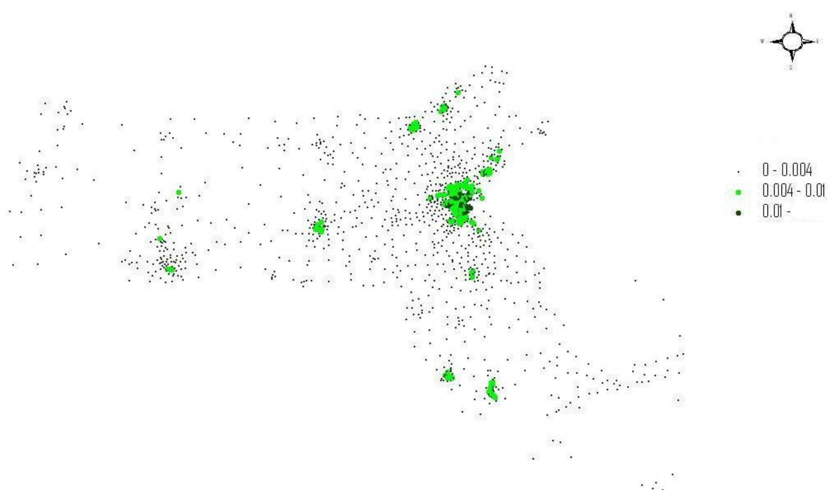


Figure 4: Distribution of population density in Massachusetts (town level) in 2000.

B Lemma 7

Lemma 7 If Assumption 2 is true then X, Y are discordant, i.e. $\tau < 0$.

Proof Let (X_1, Y_1) and (X_2, Y_2) be two independent copies of $(X, Y) \sim F_{Y|X} \times F_X$. Then

$$P((X_2 - X_1)(Y_2 - Y_1) < 0) = P\left(X_2 < X_1, Y_2 > Y_1\right) + P(X_2 > X_1, Y_2 < Y_1) \ .$$

Moreover

$$P\left(X_2 < X_1, Y_2 > Y_1\right) = \int_{\mathbb{R}} \int_{-\infty}^{x_1} \int_{\mathbb{R}} \int_{y_1}^{\infty} dF_{Y|X}(y_2 \mid x_2) dF_{Y|X}(y_1 \mid x_1) dF_X(x_2) dF_X(x_1)$$

$$= \int_{\mathbb{R}} \int_{-\infty}^{x_1} \int_{\mathbb{R}} \left[1 - F_{Y|X}(y_1 \mid x_2)\right] dF_{Y|X}(y_1 \mid x_1) dF_X(x_2) dF_X(x_1)$$

$$> \int_{\mathbb{R}} \int_{-\infty}^{x_1} \int_{\mathbb{R}} \left[1 - F_{Y|X}(y_1 \mid x_1)\right] dF_{Y|X}(y_1 \mid x_1) dF_X(x_2) dF_X(x_1) \quad \text{[by Assumption 2]}$$

$$= \int_{\mathbb{R}} \int_{-\infty}^{x_1} \left(1 - \left.\frac{F^2_{Y|X}(y_1 \mid x_1)}{2}\right|_{\mathbb{R}}\right) dF_X(x_2) dF_X(x_1)$$

$$= \frac{1}{2} \int_{\mathbb{R}} \int_{-\infty}^{x_1} dF_X(x_2) dF_X(x_1) = \frac{1}{2} \int_{\mathbb{R}} F_X(x_1) dF_X(x_1) = \frac{1}{2} \times \left.\frac{F^2_X(x_1)}{2}\right|_{\mathbb{R}} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \,.$$

By reasoning in the same manner, we obtain $P(X_2 > X_1, Y_2 < Y_1) > \frac{1}{4}$, so that

$$\pi_d = P((X_2 - X_1)(Y_2 - Y_1) < 0) > \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

and $\tau = 1 - 2\pi_d < 1 - 2 \times \frac{1}{2} = 0.$ ▮

C Check of Assumption 2 in Example 6

Let $Z_1 \sim \Gamma(c, a_1)$ and $Z_2 \sim \Gamma(c, a_2)$, with $a_1 < a_2$. Then $a_1 Z_1 \sim \Gamma(c, 1)$, $a_2 Z_2 \sim \Gamma(c, 1)$ and $P(a_1 Z_1 \leq t) = P(a_2 Z_2 \leq t)$, $\forall t$. Hence

$$P(Z_1 \leq z) = P(a_1 Z_1 \leq a_1 z) = P(a_2 Z_2 \leq a_1 z) \leq P(a_2 Z_2 \leq a_2 z) = P(Z_2 \leq z) \ \forall z \,.$$

Consider now some conditional Gamma distribution functions $\Gamma\left(c, g(x_1)\right)$ and $\Gamma\left(c, g(x_2)\right)$ where $0 < x_1 < x_2$ and $g(x)$ is a positive monotone increasing function on $(0, \infty)$. Thus, $g(x_1) < g(x_2)$ and,

$$F_{\Gamma(c,g(x_1))}(y) < F_{\Gamma(c,g(x_2))}(y), \qquad \forall y > 0 \quad \text{and } x_1 < x_2 \,. \tag{11}$$

Applying (11) to $c = \theta$ and $g(x) = \tau e^{ax}/b$ with $a > 0$, we obtain that Assumption 2 is satisfied by the Gamma model in Example 6.

D Calculation of the Kendall τ of Gamma-Gamma Model (5)

Using the model at (9) equivalent to the model at (5), we have that if $\theta = 1$, then $P(Y_{ij} > s, Y_{ih} > t)$ can be represented as the Laplace transform of a Gamma(α, α) distribution evaluated in $(s\mu_{ij}^{-1} + t\mu_{ih}^{-1})$.

Indeed:

$$P(Y_{ij} > s, Y_{ih} > t) = \mathrm{E}(P(Y_{ij} > s|w_i)P(Y_{ih} > t|w_i)) = \mathrm{E}(\mathrm{e}^{-s\mu_{ij}^{-1}w_i - t\mu_{ih}^{-1}}) \ .$$

Hence $P(Y_{ij} > s) = \left[\alpha/(\alpha + s\mu_{ij}^{-1})\right]^{\alpha}$. So $P(Y_{ij} > s, Y_{ih} > t)$ has form

$$P(Y_{ij} > s, Y_{ih} > t) = (P(Y_{ij} > s)^{-1/\alpha} + P(Y_{ih} > t)^{-1/\alpha})^{-\alpha} \ .$$

The Kendall's τ of this kind of bivariate distributions has been investigated in Example 5.4 in Nelsen (1999), where one find that $\tau = \alpha/(\alpha + 2)$.

County	Obs	Mean (Std. Dev)				(Min, Max)			
		Pop	Inc	Dist	Mix	Pop	Inc	Dist	Mix
Suffolk	4	11.345 (3.573)	21.263 (5.422)	6.74 (4.42)	0.701 (0.223)	(8.001, 16.018)	(14.628, 27.374)	(0, 10)	(0.490, 0.945)
Franklin	26	0.111 (0.164)	22.419 (4.604)	145.35 (31.77)	0.974 (0.023)	(0.009, 0.836)	(12.400, 31.891)	(29, 192)	(0.888, 1.000)
Plymouth	27	0.949 (0.997)	26.643 (6.267)	47.63 (20.83)	0.964 (0.057)	(0.135, 4.392)	(17.163, 41.703)	(5, 89)	(0.687 , 0.989)
Middlesex	54	2.948 (3.943)	33.699 (11.908)	31.87 (17.03)	0.909 (0.073)	(0.120, 18.851)	(17.557, 79.640)	(5.5, 79)	(0.664, 0.985)
Bristol	20	1.114 (1.096)	24.124 (5.451)	66.11 (16.02)	0.959 (0.037)	(0.219, 0.838)	(15.602, 41.901)	(41, 92)	(0.838, 0.988)
Berkshire	32	0.143 (0.230)	25.815 (7.913)	208.19 (10.23)	0.975 (0.019)	(0.006, 1.124)	(16.381, 50.149)	(188, 231)	(0.913 ,1.000)
Hampden	23	0.814 (1.107)	22.998 (5.216)	139.91 (27.55)	0.929 (0.121)	(0.013, 4.738)	(15.232, 38.949)	(70, 182)	(0.528 ,0.989)
Essex	34	1.909 (2.290)	30.562 (8.737)	40.21 (14.07)	0.935 (0.111)	(0.231, 10.351)	(13.360, 48.846)	(14, 62)	(0.420 ,0.988)
Hampshire	20	0.306 (0.405)	22.930 (2.957)	153.25 (21.97)	0.962 (0.045)	(0.022, 1.258)	(17.427, 29.821)	(109, 192)	(0.796 , 0.997)
Dukes	7	0.204 (0.233)	24.870 (5.387)	128.14 (9.35)	0.970 (0.022)	(0.006, 0.572)	(15.265, 31.021)	(118, 144)	(0.934,1.000)
Worcester	60	0.565 (0.717)	24.424 (5.970)	77.13 (20.48)	0.950 (0.057)	(0.022, 4.597)	(16.845, 44.310)	(42, 150)	(0.741, 0.994)
Norfolk	28	1.819 (1.690)	33.576 (10.039)	31.77 (24.31)	0.922 (0.072)	(0.363, 8.410)	(23.379, 64.899)	(8.5, 143)	(0.647 ,0.984)
Barnstable	15	0.512 (0.246)	25.359 (2.194)	127.73 (27.29)	0.963 (0.022)	(0.099, 1.023)	(22.092, 29.553)	(89, 180)	(0.896 ,0.983)
Nantucket	1	0.199	31.314	112	0.887				

Table 5: Population (Pop) densities per square mile, average income (Inc) per capita (in thousand of $) in 1999, and Proportion of white population (Mix) of each county, Source: US Bureau Census (2000), Calculus: authors. Shortest distance to Boston in km (Dist), Source:www.viamichelin.com.