On the density distribution across space: a probabilistic approach

Ilenia Epifani* - Rosella Nicolini‡

July 22, 2009

Abstract

This paper aims at providing a Bayesian parametric framework to tackle the accessibility problem across space in urban theory. Adopting continuous variables in a probabilistic setting we are able to associate with the distribution density to the Kendall's tau index and replicate the general issues related to the role of proximity in a more general context. In addition, by referring to the Beta and Gamma distribution, we are able to introduce a differentiation feature in each spatial unit without incurring in any a-priori definition of territorial units. We are also providing an empirical application of our theoretical setting to study the density distribution of the population across Massachusetts.

Keywords: Agglomerations, Bayesian inference, Distance, Gibbs sampling, Kendall's tau index, Population density.

JEL Classification: C40, R14.

1 Introduction

Empirical evidence backs the idea that the distribution of population or activities across space is not uniform. According to the accessibility concept, people show a particular interest in locating as close as possible to the central business district (CBD). In such a way they enjoy an easy access to all the amenities and other facilities they look for (see Fujita-Thisse, 2002 or Song, 1996). According to Song (1996) the concept of accessibility is very important in defining urban form

*Dipartimento di Matematica "F. Brioschi", Politecnico di Milano, Piazza Leonardo da Vinci, 32-20133 Milano (Italia), Ph. +39 0223994535. E-mail:ilenia.epifani@polimi.it.

†Insitut d'Anàlisi Econòmica - CSIC, Campus de la Universitat Autònoma de Barcelona, 08193 Bellaterra (Spain). Ph. +34935806612. E-mail: rosella.nicolini@iae.csic.es.

‡We would like to thank the participants at the 3rd Spatial Econometric World Conference for useful suggestions. The usual disclaimer applies. R. Nicolini's research is supported by a *Ramón y Cajal* contract and by the Barcelona GSE network. Financial support from research grants 2005SGR00470 and SEJ2008-01850 is acknowledged.

and function. For instance, it measures the ease of access to an economic activity from a specific location and it contributes to quantifying the market potential concept for any location (see, for instance, Glaeser, 2008). The accessibility function also introduces a heterogeneity in the space: all locations cannot be considered as equivalent from a strictly economic viewpoint. Therefore, whenever consumers display a preference for one location with respect to another, their distribution across space is not expected to be constant.

But, how could we formalize this property of the space where proximity matters? The definition of the distribution function draws the relationship between, for instance, population density and the distance from a central point (for instance the CDB) (Nairn and O'Neill, 1988). However, as listed in Song (1996), there is a wide number of possible population density functions that can be applied in scientific studies. The existence of this wide range of functions stems from the variety of accessibility functions that can be adopted. The general distance from a CBD can be measured in multiple ways. Then, in a standard urban setting, the population density function is the product of the accessibility function and an index of the population density at a single location point.

The very controversial issue is the way to (i) define a distance function, and (ii) identify the proper territorial unit to deal with the problem of the importance of proximity as an agglomeration force toward a CDB. The state-of-the-art literature generally proposes exogenous methods to introduce a distance function and, then, define the proper territorial unit.

The probabilistic approach we are proposing in this study allows to overcome some problems associated with a variety of functions due to the variety of definition of the accessibility concept. One of the key issues of the standard spatial equilibrium model (for monocentric city, for instance) is the analysis of the impact of transportation costs on the population density. Commuting entails a cost and, as a consequence, the urban structure is considered as a sort of distance-minimizing structure. Urban models dating back to Alonso-Muth-Mills setting need to look for an approximation of the distance function and, then, the costs associated with it (Glaeser, 2008). The empirical studies founded on these models also suffer from the same problem. Our approach is more general because our starting point is a simple axiomatic assumption of population preferences. Considering the population and distance functions as random continuous variables and by applying the Kendall's tau index, we are able to replicate all the previous results in a more general framework and to add new and further findings.

Furthermore, an additional advantage of this approach is to manage a differentiation feature of the space by adopting a Beta (probabilistic) distribution function, as a benchmark. In fact, the Beta distribution enhances the concept of uneven distribution of agents across space. It relaxes the assumption of uniform distribution of the population density function within the quantile and decile groups of reference. Thinking of the income distribution, the general exogenous way to define space and, hence, the uniform distribution of income across space can generate severe distortions when aggregating provincial data into regional data in order to investigate regional disparities (Chotikapanich, Rao and Tang, 2007). Instead, the Beta distribution is used for modeling events that are constrained to take place in a general interval and whose shape varies with respect to the value of parameters (McDonald and Xu, 1995).

However, the lack of a-priori information on the behavior of the density distribution across space of population, for instance, translates into the difficulty in adopting the Beta distribution as a general distribution to work with. Another candidate that can fit our scope is the Gamma distribution. One of the most interesting properties of this function is its ability to behave like other used distributions and, sometimes, Gamma distribution helps in defining which of those distributions should be adopted to model a particular process. Of course, the concern of preserving the properties of the Beta function induces us to work on the conditions that makes the adoptions of the two functions indifferent from a statistical viewpoint. Once more, the study of the Kendall tau index reveals to be the key criterion to asses the identity between the two approaches when aiming at modeling our problem.

In addition, a Gamma model is very useful if we expect an increase in the variance of the density for larger values of the mean density and, hence, a shorter distance from the CBD (see, McCullagh and Nelder 1989). Therefore, this piece of evidence suggests to exploiting the flexibility of the Gamma function to model the behaviour of the population density.

According to our approach, we can assess not only that the distribution density function decreases with the distance from the CBD but also we can connect how rapidly the density falls off with distance to the values of the parameters of the Gamma distribution. We also provide an application of our empirical strategy. We are applying an estimation method based on a "Gamma-Gamma model" to the study of the distribution of the population density of the counties in Massachusetts. By choosing Boston as CBD and, considering the distance as a probabilistic function, our likelihood function is able to replicate the distribution pattern of population density of each town in Massachusetts against its relative distance to Boston. We also run a few statistical check of robustness of our estimated parameters and, basically, our estimated results converge to the real ones.

The remainder is organized as follows. In Section 2 we describe our setting of analysis. Section 3 deals with the concept of Kendall's tau and its applications. In Section 4 we develop a Gamma-Gamma model that applied to the case of Massachusetts and Section 5 concludes. All proofs are deferred to the Appendix.

2 The setting

We aim at defining a function representing the distribution problem of a continuous of agent. We identify space with the continuous line $X \in (0, \infty)$ and the total surface of land in each location $x \in X$ is equal to one.

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

Assumption 1 bis *(Mas-Colell et al., 1995). The choice structure* $(\beta, C(.))$ *satisfies the* weak axiom of revealed preferences *if the following property holds: if for some* $X \in \beta$ *with* $(x, z) \in X$ *we have* $x \in C(X)$, *then for any* $X' \in \beta$ *with* $(x, z) \in X'$ *we must also have* $x \in C(X')$.

Without loss of generality, we can define the CBD at 0. Hence, X can be seen as the spatial distance from the CBD. Moreover, let Y be the population density in the space. The cumulative distribution function of the population density conditional to the distance $F_{Y|X}(y \mid x)$ is defined as $F_{Y|X}(y \mid x) = P(Y \leq y \mid X = x)$.

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

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

According to Assumption 2, the hypothesis we introduce regarding consumers preferences implies that it is more likely that the density of the population is lower as the distance from the CBD increases. Indeed, the inequality in Assumption 2 means that the proportion of census tracts at a distance x_1 (from the CBD) with population density less than or equal to y is no greater than the proportion of census tracts more distant (x_2), with population density less than or equal to y. In other words, under Assumption 2, large distances X from the CBD tend to be associated with small densities of population Y.

In a probabilistic setting, Assumption 2 corresponds to the classical notion of dependence introduced by Lehmann (1966) and called *Stochastically Decreasing* (SD). The Kendall's tau index measures the degree of this kind of association between X, Y, *i.e.* how much X, Y are concordant or discordant.

Definition 3 *The* Kendall's tau (τ) index *of a random vector* (X, Y) *is given by*

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

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

The Kendall's τ index assumes values in the interval $[-1.1]$ and is negative if and only if $\pi_d > 1/2$; if the Kendall's τ of X, Y is negative, then X, Y are *discordant or negative associated* random variables.

The characterization of the discordance between X and Y in terms of a restriction on the values that π_d can assume ($\pi_d > 1/2$) has an interesting intuition. The probability that either $x_1 < x_2$ is associated with $y_2 > y_1$ or $x_1 > x_2$ to $y_2 < y_1$ is greater than 1/2. This condition implies that values of (X, Y) are dissociated with a high probability, namely greater than one half.

Lemma 4 If Assumption 2 is true then X, Y are discordant, *i.e.* $\tau < 0$.

There is no a unique manner to define *a-priori* the population density distribution and density; any positive random variable can be chosen. Below, we consider some examples.

Example 5 (Log-normal Model) Let us define Y as follows:

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

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 is clearly stochastically decreasing in x if $\alpha > 0$. In Equation (2) the density of population Y is modeled as a negative exponential function of the distance from the CBD and the parameter α is the density gradient that describes how rapidly the density falls off with distance. This corresponds to the classical analysis of the accessibility problem, where, by assumption, α is assumed to be greater than zero. See, for example, the estimation function of the accessibility measure numbered 1 in the first row of Table 1 in Song (1996). Therefore, this example emphasizes that the classical log-normal regression analysis of the accessibility satisfies Assumption 2.

Example 6 (Beta-Gamma model) Let $f_{Y|X}(y \mid x)$ be a Beta density of parameters c and $ax + b$ with a, b and c all positive, *i.e.*

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{y^{c-1}(1-y)^{ax+b-1}}{B(c, ax+b)} & \text{if } 0 < y < 1 \text{ and } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $B(c, ax + b) = \int_0^1 y^{c-1}(1 - y)^{ax+b-1}$.

For any marginal density of the nonnegative random variable X, Assumption 2 is satisfied. In fact, the partial derivative of a Beta cumulative distribution function $F(w; a, \theta)$ with respect to θ

is

$$\frac{\partial F(w;a,\theta)}{\partial\theta} = \frac{\theta-1}{(B(a,\theta))^2}\int_0^1 dz_2\int_0^w dz_1(z_2-z_1)(z_1z_2)^{a-1}[(1-z_1)(1-z_2)]^{\theta-2} =$$
$$= \int_w^1 dz_2\int_0^w dz_1(z_2-z_1)(z_1z_2)^{a-1}[(1-z_1)(1-z_2)]^{\theta-2}$$

and, the last integral is positive for any $w\in(0,1)$. It follows that for any $y\in(0,1)$, any beta$(c, ax+b)$ cumulative distribution function $x\mapsto F_{Y|X}(y\mid x)$ is an increasing function of x and thus Assumption 2 is satisfied.

The conditional expected value of Y given $X=x$ is

$$\mathrm{E}(Y|X=x) = \frac{c}{c+ax+b} \qquad \forall a,b,c>0 \tag{4}$$

Notice that the right hand side in Equation (4) is equal to $1/(1+\tilde{a}+\tilde{b}X)$, with $\tilde{a}=a/c$ and $\tilde{b}=b/c$ so that the shape parameter c is not identifiable. Henceforth, to remove this problem, we assume $c=1$. If Y given $X=x$ is beta$(1,ax+b)$-distributed, then its conditional cumulative distribution is $F_{Y|X=x}(y\mid x) = 1-(1-y)^{ax+b}$, for all y in $(0,1)$ and $\mathrm{E}(Y|X=x)=1/(1+ax+b)$, with conditional variance $\mathrm{Var}(Y|X=x) = (ax+b)/[(1+ax+b)^2(2+ax+b)]$.

To complete the model for the couple (X,Y), let X be a random variable Gamma-distributed with shape parameter α and rate parameter β *i.e* its density is

$$f_X(x) = \begin{cases} \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)}\mathrm{e}^{-\beta x} & \text{if } x>0 \\ 0 & \text{otherwise}\end{cases} \tag{5}$$

We shall write $X\sim\Gamma(\alpha,\beta)$. It follows that

$$f_{X,Y}(x,y) = \begin{cases} \frac{(ax+b)\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)}(1-y)^{ax+b-1}\mathrm{e}^{-\beta x} & \text{if } 0<y<1 \text{ and } x>0 \\ 0 & \text{otherwise}\end{cases}$$

and

$$\begin{aligned}
f_Y(y) &= \int_0^\infty f_{Y|X}(y\mid x)\,f_X(x)\,dx \\
&= \frac{a\beta^\alpha(1-y)^{b-1}}{\Gamma(\alpha)}\int_0^\infty x^\alpha \mathrm{e}^{-x(\beta-a\log(1-y))}dx \;+\; \beta^\alpha b(1-y)^{b-1}\int_0^\infty \frac{x^{\alpha-1}}{\Gamma(\alpha)}\mathrm{e}^{-x(\beta-a\log(1-y))}dx \\
&= \frac{a\beta^\alpha(1-y)^{b-1}}{\Gamma(\alpha)}\times\frac{\Gamma(\alpha+1)}{[\beta-a\log(1-y)]^{\alpha+1}} + \frac{\beta^\alpha b(1-y)^{b-1}}{[\beta-a\log(1-y)]^{\alpha+1}} \\
&= \frac{(a\alpha+b)\beta^\alpha(1-y)^{b-1}}{[\beta-a\log(1-y)]^{\alpha+1}}
\end{aligned}$$

The marginal expected value of Y is given by

$$\mathrm{E}(Y) = (a\alpha+b)\int_0^\infty \frac{\beta^\alpha y(1-y)^{b-1}}{[\beta-a\log(1-y)]^{\alpha+1}}\,dx$$

which, alternatively, we can compute as

$$\mathrm{E}(Y) = \int_0^{\infty} \frac{1}{1 + \frac{a}{\beta}x + b} \times \frac{x^{\alpha-1}\mathrm{e}^{-x}}{\Gamma(\alpha)}\, dx\ . \tag{6}$$

Looking at the expression of mean $\mathrm{E}(Y)$ in Equation (6), we deduce that on average the population density depends on the ratio a/β which can be interpreted as the density gradient computed as a pure number: indeed, observe that the scale parameter β changes with changes in the scale measurement of the distance from CBD x.

As a further remark, we notice that when turning to consider the conditional mean $\mathrm{E}(Y|X = x)$ of a conditional Beta distribution, we are adding a further feature: we are assuming that the population distribution function at any distance x is not uniform. In particular, we are assuming not only that the distribution of the population is not uniform in each territorial unit, but also that its variance increases along with the distance from the CDB. Hence, we are including a further element of potential inequality across different points of the space.

Example 7 (Extended-beta Model) The beta density function on $[0, 1]$ can be easily extended to a random variable with support $[0, M]$. Hence, we can use a beta model for population density Y with values in $[0, M]$, for some $M > 0$ not necessarily equal to one. Model (3) takes the form:

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{(ax+b)(M-y)^{ax+b-1}}{M^{ax+b}} & \text{if } 0 < y < M \text{ and } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

and the corresponding cumulative distribution function is given by

$$F_{Y|X}(y \mid x) = \begin{cases} 0 & \text{if } y \leq 0 \\ 1 - (1 - \frac{y}{M})^{ax+b} & \text{if } 0 < y < M \\ 1 & y \geq M \end{cases}$$

which satisfies Assumption 2.

Example 8 (Gamma model) Suppose that conditional on $X = x$, Y is Gamma-distributed with shape θ and rate function $\theta\mathrm{e}^{ax}/b$:

$$Y|X = x \sim \Gamma\left(\theta, \theta\frac{\mathrm{e}^{ax}}{b}\right), \qquad b > 0\ . \tag{8}$$

In Appendix B, we check that Assumption 2 is satisfied if and only if $a > 0$. The conditional mean of Y given $X = x$ is $\mathrm{E}(Y|X) = b\mathrm{e}^{-ax}$ and the (conditional) coefficient of variation (CV from now on), defined as the ratio of the standard deviation to the mean, is constant and equal to $\theta^{-1/2}$, since $\mathrm{Var}(Y|X) = \mathrm{E}(Y|X)^2/\theta$.

A model with constant coefficients of variation is very useful if we expect the variance of Y to increase with its mean or smaller values of the distance X from CBD. Actually, the descriptive statistics referring to the distribution of the population density across counties in Massachusetts replicate this feature as shown in Table 6. This evidence suggests to exploiting the flexibility of the Gamma function to model the behaviour of the population density.

Using the Gamma Model (8) to deal with the accessibility is equivalent to accepting a regression model with multiplicative gamma errors for original data. In the previous Log-normal Model (2) the variance is constant and Log-normal model (2) provides an additive regression model for the logarithms of density and distance from the CBD. Conversely, in a Gamma model, density and distance from the CBD are measured on the original scale. Once more, the conditional mean $\mathrm{E}(Y|X=x)$ coincides with the notion of accessibility as distance from CDB but measured on the original scale. Anyway, if we take a log link, *i.e.* $\log \mathrm{E}(Y|X) = \log b - aX$, then Model (8) can be analyzed under the generalized linear model (GLM) setup (see, McCullagh and Nelder 1989). The parameter a can be interpreted as a measure of the density gradient describing the decreasing speed of the density against the distance, whereas parameter b describes the density at or near the CBD.

The Gamma model is very flexible. We can choose a rate function $\eta(x)$ alternative to e^{ax}/b examined here. For example, model $Y|X=x \sim \Gamma(\theta, \theta(ax+b))$, with $a>0$ satisfies Assumption 2. Our interest in the last model $Y|X \sim \Gamma(\theta, \theta(aX+b))$ comes from the close connection between its Kendall's τ and the Kendall's τ of the *Beta-Gamma model* described in the previous Example 6. The computation of the Kendall's τs is performed in Section 3.

A more articulated Gamma model fitting the description of the distribution of the population density across counties in Massachusetts will be examined in detail in Section 4.

3 Computation of the Kendall's τ

In this section we compute the Kendall's τ index for the Gamma and Beta-Gamma models described in Section 2 and analyze how they vary as functions of the parameters. This exercise is very useful to quantify to what extend the shape of the function is able to condition the decay of the density against the distance with respect to the CDB. In doing so, we are also able to trace back to the standard results usually obtained in the current literature as a particular specification of this general framework.

We start with some general remarks that turn out to be very useful first to simplify the computation of the Kendall's τ, and second to shed light what parameters effectively determine the Kendall's τ and thus influence the dependence of density population from the distance from CBD.

As a general remark, note that the value of the Kendall's τ index of a couple (X, Y) does not change under every increasing monotone (deterministic) transformations 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. In particular, $\tau(X,Y) = \tau(lX, mY)$, for all $l > 0$ and $m > 0$ and $\tau(X,Y) = \tau(X, -\log(1-Y))$ for any random variable Y with support $[0,1]$.

Example 9 (Gamma-Gamma Model) Let $\eta(x) : (0,\infty) \mapsto (0,\infty)$ be a monotone increasing function and consider the model given by: $Y|X = x \sim \Gamma(\theta, \theta\eta(x))$ and $X \sim \Gamma(\alpha,\beta), \theta, \alpha, \beta > 0$ that we shall denotes as $\Gamma(\theta,\theta\eta(x)) \times \Gamma(\alpha,\beta)$. In the light of the previous remark on θ and of the properties of the gamma distributions family, we have that 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, for all $m > 0$.

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

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

In particular, if $\eta(X) = aX$ and $m = a/\beta$ then the vectors (X,Y) and (W,Z) with $Z|W \sim \Gamma(\theta,\theta W)$ and $W \sim \Gamma(\alpha,1)$ share the same Kendall's τ. So, in a Gamma-Gamma model $\Gamma(\theta,\theta aX) \times \Gamma(\alpha,\beta)$, the Kendall's τ is independent of both the two rate parameters a and β, and depends only on the shape parameters α and θ.

On the other hand, notice that if $\eta(x) = ax + b$, then the 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 terms, if $b \neq 0$ then τ depends on the coefficient of variation CV of Y conditional on X given by $\mathrm{CV} = \theta^{-1/2}$, by shape parameter α of X, by b and by gradient a/β, computed as a pure number.

In any case, without loss of generality, in our numerical calculations of τ, we can suppose $\beta = 1$. As a matter of fact, the expression for Kendall τ of this model can only be evaluated numerically or via simulation. For example, some simplification for τ arises when α and θ are integers. For example, when $\alpha = \theta = 1$, one has

$$\tau(X,Y) = 2\int_0^\infty \int_{x_1}^\infty \int_0^\infty \int_{y_2}^\infty x_1 x_2 \mathrm{e}^{-x_1 y_1 - x_1 - x_2 y_2 - x_2}\ dy_1 dy_2 dx_2 dx_1 - 1 =$$
$$= 2\int_0^\infty \int_{x_1}^\infty \mathrm{e}^{-(x_1+x_2)} \frac{x_2}{x_1+x_2}\ dx_2\ dx_1 = 2 \times 0.375 - 1 = -0.5$$

Anyway, in general, we evaluated τ using the following simple simulation scheme. We just simulated $N = 20000$ 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*:

$$R_K = \frac{C - D}{N(N-1)} \tag{9}$$

where C is the number of concordant couples and D the number of discordant. Remember that two pairs are *concordant* if both members of a couple are larger than their 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.

Alternatively, we could evaluate τ via a simple Monte Carlo scheme based on the simulation of two independent bivariate samples $\{(X_i^{(j)}, Y_i^{(j)})\}_{i=1}^N$, for $j = 1, 2$ from the distribution of (X, Y). Hence, we obtain

$$\hat{\tau}_N = \frac{\sum_{i=1}^N \operatorname{sign}\left((X_i^{(1)} - X_i^{(2)})(Y_i^{(1)} - Y_i^{(2)})\right)}{N} .$$

The results of simulations are given in Table 1 (for fixed $\beta = 1$).

One observes in Table 1 that for fixed parameter b and the conditional coefficient of variation CV of Y given X, the negative dependence decreases as α increases (the Kendall's τ approaches 0 from bellow as $\alpha \to \infty$). On the other hand, given the shape parameter α of the gamma distance X, the negative dependence decreases with the conditional coefficient of variation $\mathrm{CV} = \theta^{-1/2}$ of the conditional gamma distribution of Y given X. Actually, larger values of CV imply larger dispersion for Y, whereas larger values of α imply larger variance for X and so, in both these cases, smaller dependence between X and Y. Finally, given $\mathrm{CV} = \theta^{-1/2}$, $b \neq 0$ and α, the dependence between X and Y increases as gradient a increases and is the more relevant the bigger are b and θ[1].

Thinking of τ as the degree of dissociation in the location choices of agents, the negative dependence reinforces proportionally to shape θ, *i.e.* the distribution density-polarizes. Put differently, whenever the CBD enlarges, the preferences of agents coincides because they all want to settle close to the CBD (for any kind of function of distance shaping the space). Beside, still for any function of distance, agents' preferences (with respect to a possible location in proximity of the CBD) become more blurred when α increases (the shape parameter of the distance function), because the distance function becomes polarized up to make this spatial dimension disappear. An extreme point would be reducing the space to one single point.

Example 10 (Gamma model. Continued) Let $Y|X \sim \Gamma\left(\theta, \theta\frac{\mathrm{e}^{aX}}{b}\right)$ with $b > 1, a > 0$ and $X \sim \Gamma(\alpha, \beta)$. One can verify that

$$\tau(X, Y) = \tau(\beta X, Y/b) = \tau(W, Z),$$

where $W \sim \Gamma(\alpha, 1)$ and $Z|W \sim \Gamma(\theta, \theta\mathrm{e}^{aW/\beta})$. In other terms, the value of τ depends only on the coefficient of variation $\mathrm{CV}(Y|X) = \theta^{-1/2}$,the shape α of X and the ratio a/β.

We evaluated Kendall's τ of the Gamma-Gamma Model (8) by using the simulation scheme described in Example 9 and empirical Formula (9). A summary of the values of the empirical R_K for $\beta = 1$ is in Tables 2.

[1]Note that 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 1: Kendall's τ for the Gamma-Gamma model in Example 9 with $\mathrm{E}(Y|X) = 1/(aX + b)$, coefficient of variation $\mathrm{CV} = 1/\theta^{-1/2}$ and $\beta = 1$.

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

Table 2: Kendall's τ for Gamma-Gamma models with $\mathrm{E}(Y|X) = e^{-aX}$ and $\beta = 1$. See Example 10 and Equation (8).

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

As in the previous case, the degree of association of the preferences of agents, measured by τ, reduces as long as the distances polarizes (or degenerates) in one point.

Example 11 (Beta-Gamma Model) Here we consider the following model: $Y|X \sim \mathrm{beta}(c, aX+b)$ and $X \sim \Gamma(\alpha, \beta)$, with $a, c, \alpha, \beta > 0$; we shall denotes this model as the Beta-Gamma(c, a, b, α, β) model. To compute the Kendall's τ of a beta-gamma(c, a, b, α, β) model we can use the same calculations as executed for the Gamma-Gamma model of Example 9. As already discussed in Example 6, to remove the identifiability problem we assume $c = 1$ and face with a Beta-Gamma$(1, a, b, \alpha, \beta)$ model that has a conditional mean $\mathrm{E}(Y|X) = 1/(1 + aX + b)$. Second, since the Kendall's τ index of a couple (X, Y) is invariant under every increasing monotone transformations of X or Y or X and Y, then the Kendall $\tau(X, Y)$ of a Beta-Gamma$(1, a, b, \alpha, \beta)$ model is equal to $\tau(X, Z)$, with $Z - \log(1 - Y)$. On the other hand, conditionally on X, the random variable Z is Gamma-distributed with shape parameter 1 and rate function: $aX + b$ ($i.e$ $Z \mid X \sim \Gamma(1, aX + b)$). So, we conclude that the Kendall's τ of a Beta-Gamma$(1, a, b, \alpha, \beta)$ model coincides with Kendall's τ of Gamma-Gamma model $\Gamma(1, ax + b) \times \Gamma(\alpha, \beta)$. Evaluations of Kendall's τ are summarized in Table 3, obtained by extracting the rows of Table 1 with the coefficient of variation CV = 1.

On one hand, this example reiterates the results associated with the polarization of the distance function. On the other, for a given (not polarized) distance function, as much proximity to the CDB matters for agents (by parameter a) as τ will represent a good measure of their preferences in concentrating as closely as possible to the CBD.

Table 3: Kendall's τ for the Beta-Gamma model in Example 11 with $\mathrm{E}(Y|X) = 1/(1 + aX + b)$ and $\beta = 1$.

		α			
a	b	1	10	50	100
0.5	0	-0.499	-0.175	-0.079	-0.063
0.5	1.0	-0.147	-0.148	-0.078	-0.067
0.5	10	-0.018	-0.057	-0.063	-0.058
1.0	0	-0.504	-0.175	-0.071	-0.058
1.0	1.0	-0.227	-0.160	-0.074	-0.058
1.0	10	-0.047	-0.098	-0.065	-0.052
10.0	0	-0.507	-0.174	-0.072	-0.052
10.0	1.0	-0.433	-0.175	-0.072	-0.058
10.0	10	-0.233	-0.162	-0.072	-0.053

4 A case study: the Gamma-Gamma Model

In order to illustrate the properties of our framework, we are proposing an empirical application to the case of Massachusetts.

Data are taken from US Bureau Census and refer to the year 2000. We are considering all the towns belonging to the state (351) grouped by county (14). By comparing descriptive statistics presented in Table 6,first of all it is easy to recognize the association between lower population density and a greater distance from Boston (the capital of the state). In Table 6, the distance is understood as the shortest distance of each county from Boston. Moreover, if we look at the empirical means and standard deviations of the population density (and housing density) across counties in Massachusetts, we note an increase in the standard deviation in correspondence of larger values of the mean. So, it is quite evident that there exists a clear trend replicating the properties studied for a Gamma model with constant coefficients of variation: high density's variance in correspondence to high average density. This property makes the Gamma-Gamma model suitable for the adoption.

Empirical evidence discussed above does suggests that the geographic CDB in Massachusetts may be the capital (Boston); therefore we select it as CBD and we organize the data by considering the k counties of the state. For $i = 1, \ldots, k$ let n_i be the number of cities and Z_i the size of free land (namely water areas) in ith county. Predictor Z_i is a kind of measure for the proportion of rural land in a county. Furthermore, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, k$, let Y_{ij} be the density of population of the jth city within the ith county and D_{ij} its distance from the CBD.

In our model we introduce some state variables $X_1, \ldots, X_k$ and adopt the following two-stage Gamma-Gamma model:

$$\begin{aligned} Y_{ij}|X_i &\sim \text{Gamma}(\theta, \theta X_i \mathrm{e}^{a_i D_{ij}}) \\ X_i &\sim \text{Gamma}(\alpha, \alpha B_i) \end{aligned} \tag{10}$$

with $a_i = \beta_2 + \beta_3 Z_i$ and $B_i = \exp\{\beta_0 + \beta_1 Z_i\}$. All the state variables X_i are independent, whereas the densities of population Y_{ij} are assumed to be independent across counties and dependent within.

Our framework provides a technique to estimate the density of population of a city using a) the CBD distance D_{ij} as a local covariate variable, b) the rural degree Z_i as a global covariate and c) an unobserved state variable X_i.

In some sense X_i represents all the predictors of the population density, either observable or not observable or neglected, common to all the cities in the ith county. According to the contents of the last World Development Report (World Bank, 2008) and Ramcharan (2009), the economic concept of distance is something more than the Euclidean (physical) distance. In economics, distance refers to the ease or difficulty for goods, services, labor, capital, information or ideas to move across the space. Then, the cultural proximity or the quality of infrastructure can affect the economic distance between two places, even if the Euclidean distance between them is identical. In this exercise we aim at recovering this wider idea of distance, and this is the reason to look for a bunch of predictors (for population density) in addition to the physical distance. Nevertheless, by ideally ranking the different factors composing the measure of the distance, the Euclidean component is always considered as the most relevant one. Furthermore, our model is an attempt to take into account of spatial dependence in each country. In our exercise, we are assuming that the land organizational structure of each county is independent of that of the others, but towns belonging to the same county are characterized by very similar features. For instance, it is likely that citizens are submitted to local laws of own county that can be different from that of another, as well as each county my have peculiar natural endowments that others do not have, (*i.e.* Dukes is an island). Put differently, towns belonging to a same county share some common features that can be associated to fixed effects recorded in Z_is and random effects recorded in unobserved covariates X_is. For this kind of reasons, we think is sensible to model the population densities of different counties as independent random variables, and the spatial dependence between densities within counties via a state variable 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:

$$\mathrm{E}(Y_{ij}) = \frac{\alpha}{\alpha - 1} \mathrm{e}^{\beta_0 + \beta_1 Z_i - \beta_2 D_{ij} - \beta_3 Z_i D_{ij}} \tag{11}$$

whereas for $\alpha > 2$ we obtain:

$$\mathrm{Var}(Y_{ij}) = \frac{\alpha - 1 + \theta}{(\alpha - 2)\theta}(\mathrm{E}(Y_{ij}))^2 \ , \tag{12}$$

$$\mathrm{Cov}(Y_{ij}Y_{lj}) = \frac{E(Y_{ij})E(Y_{lj})}{\alpha - 2}\mathbf{1}(i = l) \ , \tag{13}$$

$$\rho(Y_{ij}Y_{ih}) = \frac{\theta}{\theta + \alpha - 1} \ , \tag{14}$$

since:

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

We read in Equation (11) that the unconditional expectation of Y_{ij} describes a log-linear regression model that includes the local and global predictors and an interaction term. The variance of Y_{ij} is quadratic in the mean (see (12)) and the correlation between the densities of the populations of two cities of the same county is always positive, since $\rho(Y_{ij}Y_{ih}) > 0$. The shape parameters θ (of the conditional law of Y_{ij} given X_i) and α (of X_{ij}) measure the intensity of the relationship between Y_{ij}, Y_{ih} (within counties). In particular, as one can see in Equation (14), the larger α, the more the state variables X_{ij} concentrate around 1; the independence of the Y_{ij} within counties is obtained for $\alpha \to 0$. Similarly, small values of α record a strong positive relationship of densities among cities in the same county, but stronger heterogeneity among the counties. Moreover, for the α value being equal, the larger θ is, the less Y_{ij}'s are concentrated and, the bigger the dependence between Y_{ij}, Y_{ih} is. Put differently, when distance does not have a discriminating impact on population distribution within each county, i.e X_i concentrates around 1, other kinds of factors have to be considered as potential discriminatory devices (here represented by the parameters shaping the distribution function). Instead, when those factors are somewhat identical across cities, within the same county, it is less likely to differentiate the population density of one city from another.

Alternately, one can measure the dependence between the densities of populations of two cities in a county, using the Kendall's τ coefficient. In the Appendix we prove that for $\theta = 1$, $\tau(Y_{ij}, Y_{ih})$ is given by

$$\tau(Y_{ij}, Y_{ih}) = \frac{2}{2 + \alpha}$$

As $\rho(Y_{ij}, Y_{ih})$ in (14), the Kendall's τ depends only on parameter α. Moreover, for $\alpha \to \infty$, both ρ and τ go to zero as $1/\alpha$; but ρ approaches 0 faster than τ. Nevertheless, in Model (10) the dependence between Y_{ij}, Y_{ih} is not linear and, then, it could not be detected only by ρ.

Furthermore, the Kendall's τ allows for detecting the dependence even if α is less than 2, whereas some of the mean values of $Y_{ij}, Y_{i,h}$ involved in Equations (11) -(14) do not exist for $\alpha \leq 2$.

4.1 Likelihood specifications of the Gamma-Gamma model

The next step is to define a suitable specification in order to estimate the population density in accordance with the statistical framework we developed. Let now $(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z})$ be the collection of the triplets (Y_{ij}, D_{ij}, Z_i) observed for every $j = 1, \ldots, n_i$ and $i = 1, \ldots, k$ and let $\boldsymbol{\beta} = (\beta_0, \beta_2, \beta_2, \beta_3)$ be the vector of the regression parameters. The likelihood function of the parameters $(\boldsymbol{\beta}, \theta, \alpha)$ based on observed data $(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z})$ can be obtained by integrating the conditional joint probability densities of $Y_{ij}|X_i$ over all the random state variables X_i's. We have

$$
\begin{aligned}
L(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha) &= \prod_{i=1}^{k} \int_0^\infty \prod_{j=1}^{n_i} \Gamma(\theta, \theta x_i \mathrm{e}^{a_i D_{ij}})(y_{ij}) \times \Gamma(\alpha, \alpha B_i)(x_i)\ dx_i \\
&= \left(\prod_{i,j} y_{ij} \right)^{\theta - 1} \times \frac{\theta^{n\theta} \alpha^{k\alpha} \prod_{i=1}^{k} \Gamma(n_i \theta + \alpha)}{\Gamma(\theta)^n \Gamma(\alpha)^k} \times \prod_{i=1}^{k} B_i^{\alpha} \left(\alpha B_i + \theta \sum_j y_{i,j} \mathrm{e}^{a_i D_{i,j}} \right)^{-n\theta - \alpha}
\end{aligned} \tag{15}
$$

with $n = \sum_{i=1}^{k} n_i$. For $\theta \neq 1$, the final form of the likelihood $L(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha)$ from (15) is too complicated to work with. In order to solve this problem, let us consider the following model

$$
\begin{cases}
Y_{i,j}|w_i \sim \text{Gamma}(\theta, \theta w_i/\mu_{ij}) \text{ independent for all } j \\
w_i \sim \text{Gamma}(\alpha, \alpha) \text{ independent for all } i \\
\mu_{ij} = \mathrm{e}^{\beta_0 + \beta_1 Z_i - \beta_2 D_{ij} - \beta_3 Z_i D_{ij}}
\end{cases} \tag{16}
$$

One can realize that Models (10) and (16) give rise to the same likelihood $L(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha)$ in (15). However, the random factors X_i and w_i are unobservable so that the two models are not distinguishable and hence every estimate of parameters $(\boldsymbol{\beta}, \theta, \alpha)$ obtained using the likelihood function should be the same for Models (10) and (16). Due to its simplicity, we use Model (16) to develop an estimation procedure of $(\boldsymbol{\beta}, \theta, \alpha)$.

Remark 12 Model (16) is not only a convenient expedient to handle a complicated likelihood. Actually, when $\theta = 1$, Model (16) is an example of an exponential regression model with scale parameter e^{β_0} and gamma shared frailties $w_1, \ldots, w_k$, also known as *gamma county random effects*. Shared frailties models have been applied in multivariate survival analysis and extensively studied in literature. Several procedures of statistical inference, both frequentist and Bayesian, have been developed. See, for example Chapter 8 in Hougaard 2000 for a review on the frequentist techniques and, for instance, Sahu, Dey, Aslanidou and Sinha (1997) for a Bayesian modeling. On the other hand, if $\theta \neq 1$, our model does not fall into the class of parametric shared frailties models, where, typically, the conditional hazard function of Y_{ij} given w_i defined as

$$h_{ij}(y) = \frac{f_{Y_{ij}|w_i}(y|w_i)}{P(Y_{ij} > t|w_i)}$$

is modeled as the product of three terms: a "baseline" hazard function h_0, a frailty term w_i and an exponential regression model $\mathrm{e}^{\boldsymbol{\beta}\mathbf{X}}$, *i.e* $h_{ij}(y) = h_0(y)w_i\mathrm{e}^{\boldsymbol{\beta}\mathbf{X}}$. Conversely, our gamma hazard function cannot be reduced to this form.

4.2 Prior specifications and Bayesian estimation

Here we develop a Bayesian technique for estimating parameters $\boldsymbol{\beta}, \alpha, \theta$. In a Bayesian perspective, the unknown parameters $\boldsymbol{\beta}, \alpha, \theta$ are understood as random variables with a *prior* joint distribution, say π and, the statistical problem consists in updating π by computing a *posterior* joint conditional probability of $\boldsymbol{\beta}, \alpha, \theta$, given data $\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}$. Then the posterior joint distribution is summarized in a simple way, typically by means of posterior means, giving rise to a point estimate of $\boldsymbol{\beta}, \alpha, \theta$. Moreover, the associated standard errors for the posterior means of $\boldsymbol{\beta}, \alpha, \theta$ are computed. We find out that both the joint and the marginal posterior distributions of $\boldsymbol{\beta}, \alpha, \theta$ does not have a closed form. So we need to use some Markov Chains Monte Carlo (MCMC) algorithms in summarizing that. In particular, we will use a Gibbs sampling scheme.

Let $\boldsymbol{w} = (w_1, \ldots, w_k)$ be the vector of the unobservable "frailties" and let us consider $(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z})$ as "complete data". Instead of $L(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha)$ from (15), let us work with the "complete" likelihood $L(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha)$ given by

$$L(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha) \;=\; \frac{\theta^{n\theta}\alpha^{k\alpha}}{\Gamma(\theta)^n\Gamma(\alpha)^k} \prod_{i,j} \left(\mu_{ij}^{\theta} y_{ij}^{\theta-1}\right) \mathrm{e}^{-\sum_{i=1}^{k} w_i(\theta\sum_j \mu_{ij}y_{ij}+\alpha)} \prod_i w_i^{n_i\theta+\alpha-1} \tag{17}$$

and handle unknown frailties $\boldsymbol{w}$ as unknown parameters to estimate.

As regards the prior, we chose "non-informative" priors for $\boldsymbol{\beta}, \alpha, \theta$ to represent our vague prior knowledge of them. A priori $\boldsymbol{\beta}, \alpha$ and θ are assumed to be independent. The regression parameters in $\boldsymbol{\beta}$ are a priori independent normal random variables with large variances:

$$\beta_k \sim \text{ Normal}(0, 10\,000)$$

and, the prior distribution for the shape parameter α is Gamma with mean 1 and large variance, *i.e.*

$$\alpha \sim \text{Gamma}(0.01, 0.01).$$

Analogously, θ is assigned Gamma(ν, ν) prior with a small ν (in our computation, $\nu = 0.01$).

Using complete likelihood $L(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}; \boldsymbol{\beta}, \theta, \alpha)$ in (17) and the priors specified above, we obtain the following full conditional distributions of each parameter given all the others and the data:

Let $\pi_{\boldsymbol{\beta}}, \pi_\theta, \pi_\alpha$ denote the prior densities of $\boldsymbol{\beta}, \alpha$ and θ, respectively and, let us denote the set of data $\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}$ by "$\boldsymbol{Data}$". Hence

1. conditional on $\boldsymbol{\beta}, \theta, \alpha$ and $\boldsymbol{Data}$, frailties terms $w_1, \dots, w_k$ are independent and Gamma-distributed:

$$w_i \sim \text{Gamma}\Big(n_i\theta + \alpha, \theta \sum_j \mu_{ij} y_{ij} + \alpha\Big) \ ;$$

2. The full conditional distribution $\pi_\alpha(\cdot|\boldsymbol{\beta}, \theta, \boldsymbol{w}, \boldsymbol{Data})$ of α, given $\boldsymbol{\beta}, \theta, \boldsymbol{w}$ and $\boldsymbol{Data}$, is proportional to

$$\frac{\alpha^{\alpha k}}{\Gamma(\alpha)^k} \mathrm{e}^{-\alpha \sum_{i=1}^k w_i} \left(\prod_i w_i\right)^{\alpha - 1} \pi_\alpha(\cdot)$$

3. The full conditional distribution $\pi_\theta(\cdot|\boldsymbol{\beta}, \alpha, \boldsymbol{w}, \boldsymbol{Data})$ of θ, given $\boldsymbol{\beta}, \alpha, \boldsymbol{w}$ and $\boldsymbol{Data}$, is proportional to

$$\frac{\theta^{n\theta}}{\Gamma(\theta)^n} \mathrm{e}^{-\theta \sum_{i,j} w_i \mu_{ij} y_{ij}} \prod_{i,j} (w_i \mu_{ij} y_{ij})^{\theta - 1} \ \pi_\theta(\cdot) \ .$$

4. The full conditional distribution $\pi_{\boldsymbol{\beta}}(\cdot|\theta, \alpha, \boldsymbol{w}, \boldsymbol{Data})$ of $\boldsymbol{\beta}$, given $\theta, \alpha, \boldsymbol{w}$ and $\boldsymbol{Data}$, is proportional to

$$\mathrm{e}^{\boldsymbol{\beta}' \theta \mathbf{c} - \theta \sum_{i,j} w_i y_{ij} \mathrm{e}^{\boldsymbol{\beta}' \mathbf{c}_{ij}}} \ \pi_{\boldsymbol{\beta}}(\cdot) \ ,$$

where $\mathbf{c}_{ij} = (1, Z_i, -D_{ij}, -Z_i D_{ij})$ and $\mathbf{c} = \sum_{i,j} \mathbf{c}_{ij}$

We can now sample frailties $\boldsymbol{w}$ and $\boldsymbol{\beta}, \alpha, \theta$ alternately sampling $\boldsymbol{w}$ and $\boldsymbol{\beta}, \alpha, \theta$ from their full conditional probability distribution as follows: given starting values $\boldsymbol{w}^{(0)}, \boldsymbol{\beta}^{(0)}, \alpha^{(0)}, \theta^{(0)}$ repeat

$$\begin{aligned}
w_i^{(l)} &\sim \text{Gamma}\Big(n_i \theta^{(l-1)} + \alpha^{(l-1)}, \theta \sum_j \mu_{ij}^{(l-1)} y_{ij} + \alpha\Big), \qquad \text{for } i = 1, \dots, k \\
\alpha^{(l)} &\sim \pi_\alpha(\cdot|\boldsymbol{\beta}^{(l-1)}, \theta^{(l-1)}, \boldsymbol{w}^{(l)}, \boldsymbol{Data}) \\
\theta^{(l)} &\sim \pi_\theta(\cdot|\boldsymbol{\beta}^{(l-1)}, \alpha^{(l)}, \boldsymbol{w}^{(l)}, \boldsymbol{Data}) \\
\boldsymbol{\beta}^{(l)} &\sim \pi_{\boldsymbol{\beta}}(\cdot|\theta^{(l)}, \alpha^{(l)}, \boldsymbol{w}^{(l)}, \boldsymbol{Data}) \ .
\end{aligned}$$

The sample of $\boldsymbol{w}, \boldsymbol{\beta}, \alpha, \theta$ so obtained, after a burn-in period, can be considered as a sample from the joint posterior distribution of the parameters. Anyway, a complete description of the Gibbs sampling is beyond the scope of this paper, further details can be found for example in Casella and and George (1992).

For carrying out the Gibbs sampler, we use JAGS (Just Another Gibbs Sampler) software package by Plummer 2009. JAGS is designed to work closely with the R package where all statistical computations and graphics are performed.

4.3 Results for the Massachusetts case study

We run few simulations for our specific Massachusetts case study. In order to deal with tractable values, we adopt the following strategies. We normalize the distance D_{ij} of each single town from

Table 4: Estimation with real data and without interaction ($\beta_3 = 0$).

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat
α	2.9	1.3	1.0	1.9	2.6	3.5	6.2	1.0
β_0	1.5	0.3	0.9	1.3	1.5	1.7	2.1	1.0
β_1	-0.4	0.8	-2.0	-0.8	-0.3	0.1	1.0	1.0
β_2	4.8	0.5	3.9	4.4	4.8	5.1	5.8	1.0
θ	1.1	0.1	1.0	1.1	1.1	1.2	1.3	1.0

the CBD (here Boston city) to a value belonging to interval $(0, 1)$. The measure of distance we are applying is:

$$\tilde{D}_{ij} = \frac{D_{ij} - \min D}{\max D - \min D}$$

where D_{ij} is the direct measure of the distance of each town to Boston, and $\min D, \max D$ are the minimum and maximum values of observed distances in our sample. In the similar manner we define

$$\tilde{Z}_i = \frac{Z_i - \min Z}{\max Z - \min Z} \ ,$$

where Z_i is the fixed effect given by the size of the free land in county i and $\min Z, \max Z$ are the minimum and maximum values of observed Z_i, respectively.

By using the package JAGS, 10 000 iterations for three chains were run for the unknown parameters $\boldsymbol{\beta}, \alpha, \theta$ and frailties $w_1, \dots, w_{14}$, and the first half was discarded as burn-in. After burn-in, one out of every 5th simulated values were kept for posterior analysis, for a total of 3000 simulations saved. Only results for effective parameters $\boldsymbol{\beta}, \alpha, \theta$ are here included. Table 4 presents the estimation results for a Model (16) without interaction (*i.e.* $\beta_3 = 0$) and Table 5 that for a Model (16) with interaction. These tables summarize the posterior mean, standard deviation and sample quantiles (2.5th, 25th, 50th, 75th and 97.5th) of parameters $\boldsymbol{\beta}, \alpha, \theta$, given $\boldsymbol{Data}$. For each parameter, the last column of Table 5 provides an estimate "Rhat" of the Gelman-Rubin potential scale reduction factor diagnostic which measures the convergence of the Gibbs sequences to the posterior distribution. In short, Rhat compares the between and within variances of the three simulated chains and, at convergence, Rhat=1. See Gelman and Rubin (1992).

To assess the goodness of fit of our model, we follow the guidelines on the Bayesian model checking contained, for example, in Albert (2009). We simulate draws of the posterior predictive density $f(y|\boldsymbol{Data})$ of the density of population for each town in Massachusetts and summarize that by the 5th and 95th quantiles. Hence we graph that as line plots in Figure 1, where the

Table 5: Estimation with real data and with interaction term

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat
α	3.3	1.7	1.1	2.1	3.0	4.2	7.7	1.0
β_0	1.2	0.4	0.6	1.0	1.2	1.5	2.0	1.0
β_1	1.1	1.7	-2.9	0.1	1.2	2.3	4.2	1.0
β_2	4.4	0.7	3.2	3.9	4.4	4.8	5.8	1.0
β_3	3.0	3.3	-4.3	1.0	3.2	5.4	9.1	1.0
θ	1.1	0.1	1.0	1.1	1.1	1.2	1.3	1.0

observed densities y_{ij} are placed as solid dots. If the actual y_{ij} turns out to be in the tail of this distribution, that indicates y_{ij} is an outlier and the model does not fit.

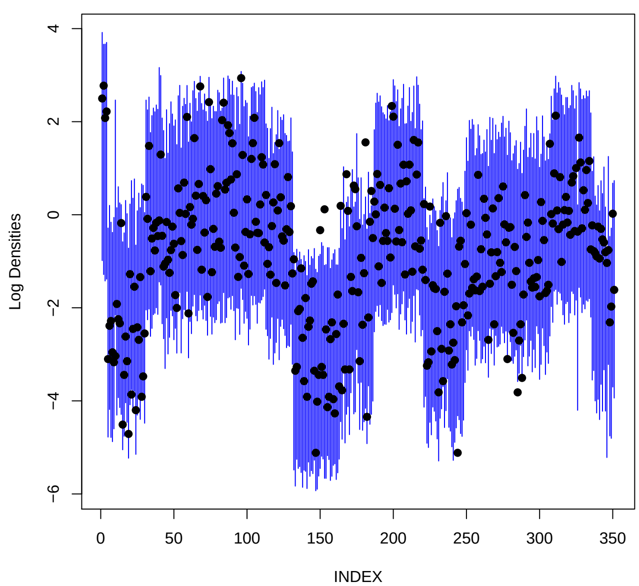

Figure 1: Posterior predictive distributions of the densities population with actual densities y_{ij} denoted by solid dots.

We note in Figure 1 that "almost all" the actual values y_{ij} are consistent with the corresponding predictive distributions. There are five points below the 5th quantile: points indexed by $10, 21$ and 28 corresponding to three towns in Franklin county, 244 which is a town in Dukes county and 285 in Worcester. A further ten points exceed the 95th quantile: town numbered 122 in Bristol county, towns 150 and 153 in Berkshire, towns $164, 168, 173, 174, 181, 185$ in Hampden, and town

225 in Hampshire.

Furthermore, we use a Q-Q plot of the empirical quantiles of actual y_{ij} versus the quantiles of 351 values (one-per town) of population density simulated according to the Model (16) with the Bayesian estimates of the parameters, provided on the first column of Table 5, plugged in. We plot out results in Figures 2 and 3.

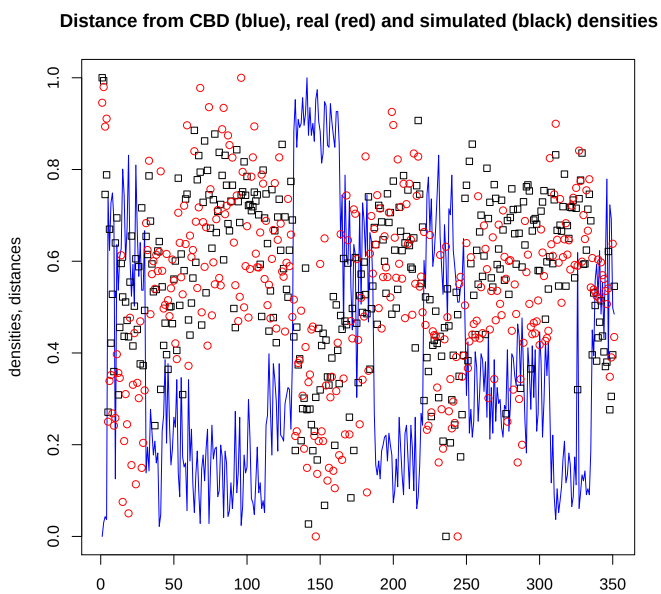

Figure 2: Plot of the simulated densities, denoted by (black) squares, real densities denoted by (red) circles and normalized distance denoted by (blue) line.

It is easy to detect that population density is generally reducing in the presence of a large distance from Boston. Moreover, the simulated and real data generally behave in a quite similar way. Anyway, the Q-Q plot of the sample quantiles of simulated densities versus that those of real densities in Figure 3 shows the tails of the simulated densities are slightly longer than that of real densities. On the other hand we have already noticed the presence of a few outliers belonging to Franklin, Dukes, Worcester, Bristol, Berkshire, Hampden and Hampshire counties. In the plots of the densities of population versus distance from Boston in Figures 4 and Figures 5 we indicated these outliers by solid points. In particular, Figure 5 shows that a negative dependence between distance from Boston and density of population seems doubtful for Bristol, Berkshire and Hampden counties. One reason we can put forward to justify this behavior is the fact that all these counties are border areas, so that it is very likely that the attractiveness of Boston may be smoothed by the degree of attractiveness of others towns or state capital such as Providence.

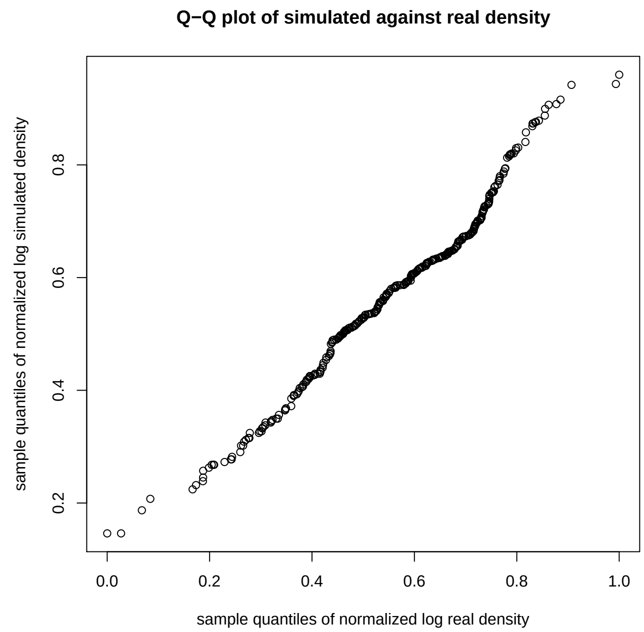

Figure 3: Q-Q plot of the sample quantiles of simulated densities versus that of real densities.

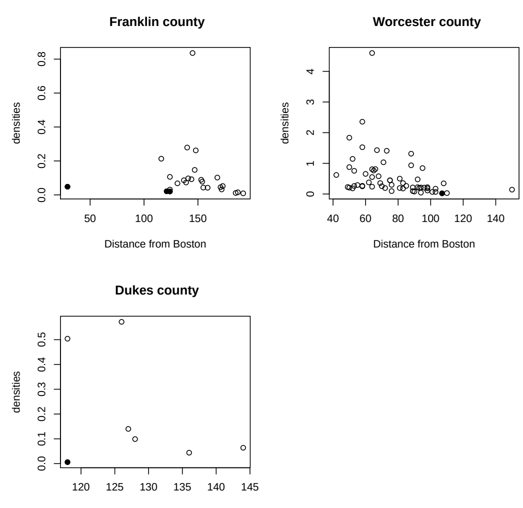

Figure 4: Plots of the log density population versus distance from Boston, with "down-outliers" points indicated by solid dots.

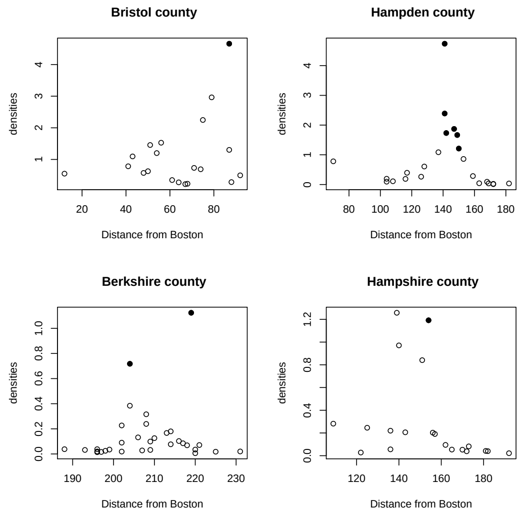

Figure 5: Plots of the log density population versus distance from Boston, with "up-outliers" points indicated by solid dots.

5 Concluding remarks

Our study proposed a probabilistic approach to estimate the distribution density in the proximity of CBD. Our framework is very general since we are following an axiomatic approach. In order to achieve our scope, we are adopting the idea of Kendall's τ index to enhance the importance of the individual preferences for settling close to the CBD. The empirical strategy we are adopting pegs on the statistical property of the Gamma function, and those properties allow to take into account the heterogeneity of the space as claimed in spatial theory. We are also proposing a preliminary empirical exercise to test the goodness of our estimation strategy. The case of Massachusetts s revealed to be a good benchmark test. The organization of the space seems to find a core in Boston city. According to the data available at hand, our predictions on the distribution of population density (against distance) across space replicate the original data well enough.

Future extensions of the study should target to extend the empirical exercise to other case studies as well as thinking about a possible extension to a multicenter spatial configuration instead of a monocentric one, as well as testing this estimation strategy for other sample of data.

Appendix

A Proof Lemma 5

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

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

Moreover

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

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

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

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

$$= \frac{1}{2} \int\limits_{\mathbb{R}} \int\limits_{-\infty}^{x_1} dF_X(x_2) dF_X(x_1)$$

$$= \frac{1}{2} \int\limits_{\mathbb{R}} F_X(x_1) dF_X(x_1)$$

$$= \frac{1}{2} \times \left.\frac{F^2_X(x_1)}{2}\right|_{\mathbb{R}} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

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

B Check of Assumption 2 in Example 7

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

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

Let us now consider some conditional Gamma distribution functions $\Gamma(c, g(x_1))$ and $\Gamma(c, g(x_2))$ 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,

$$F_{\Gamma(c,g(x_1))}(y) < F_{\Gamma(c,g(x_2))}(y), \qquad \forall y > 0 \quad \text{and } x_1 < x_2 \ . \tag{18}$$

By applying Equation (18) to $c = \theta$ and $g(x) = \tau \mathrm{e}^{ax}/b$ with $a > 0$, we obtain that Assumption 2 is satisfied by the Gamma model in Example 8. ▮

C Kendall τ of the Gamma-Gamma Model

Using Model (16), equivalent to Model (10), 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

$$P(Y_{ij} > s) = \left(\frac{\alpha}{\alpha + s\mu_{ij}^{-1}}\right)^{\alpha}$$

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 is investigated in Example 5.4 in Nelsen (1999), where one find that $\tau = \alpha/(\alpha + 2)$.

D Map of Massachussets

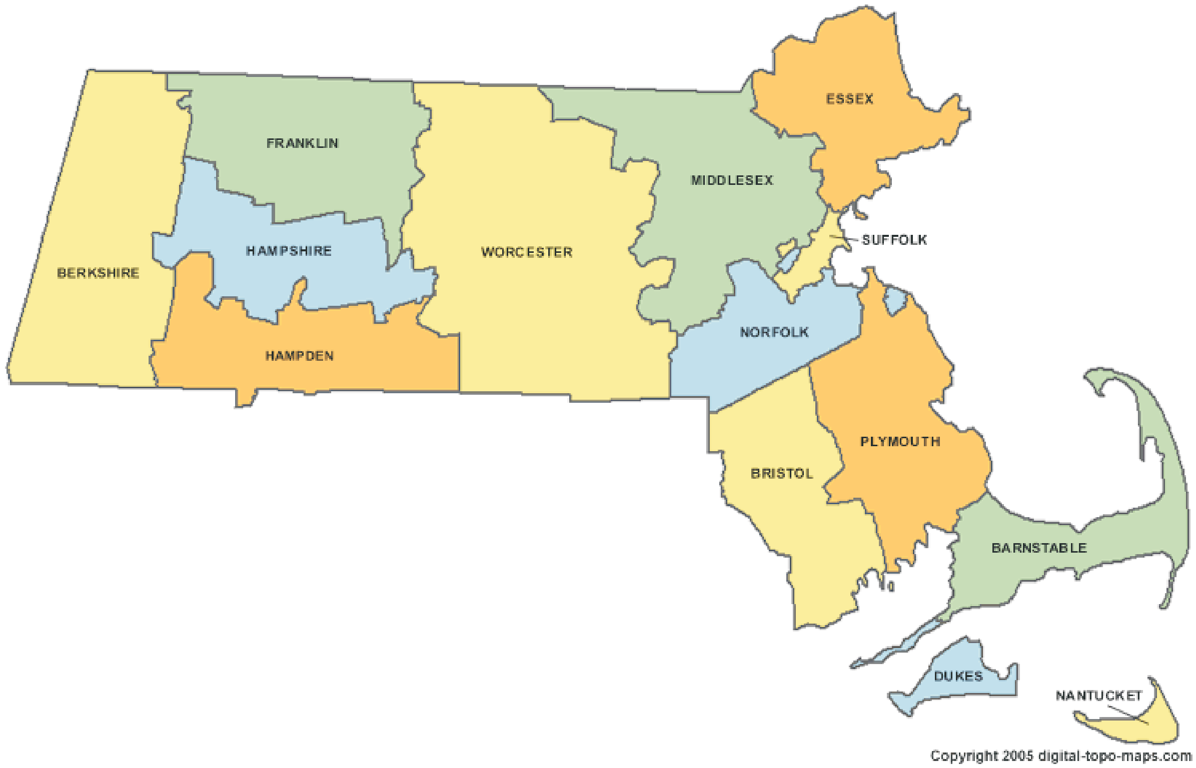

Figure 6: Map of Massachussets. Source: US Census Bureau

E Descriptive statistics about counties in Massachusetts

		Mean			Std. Dev			Min			max		
County	Obs	Pop	Hous	Dist	Pop	Hous	Dist	Pop	Hous	Dist	Pop	Hous	Dist
Suffolk	4	11.345	4.576	6.74	3.573	1.022	4.42	8.001	3.415	0	16.018	5.633	10
Franklin	26	0.111	0.050	145.35	0.164	0.075	31.77	0.009	0.006	29	0.836	0.382	192
Plymouth	27	0.949	0.372	47.63	0.997	0.417	20.83	0.135	0.048	5	4.392	1.771	89
Middlesex	54	2.948	1.205	31.87	3.943	1.693	17.03	0.120	0.042	5.5	18.851	7.902	79
Bristol	20	1.114	0.452	66.11	1.096	0.493	16.02	0.219	0.077	41	4.660	2.063	92
Berkshire	32	0.143	0.071	208.19	0.230	0.107	10.23	0.006	0.006	188	1.124	0.525	231
Hampden	23	0.814	0.332	139.91	1.107	0.453	27.55	0.013	0.010	70	4.738	1.906	182
Essex	34	1.909	0.767	40.21	2.290	0.886	14.07	0.231	0.102	14	10.351	3.678	62
Hampshire	20	0.306	0.120	153.25	0.405	0.153	21.97	0.022	0.011	109	1.258	0.528	192
Dukes	7	0.204	0.192	128.14	0.233	0.195	9.35	0.006	0.016	118	0.572	0.518	144
Worcester	60	0.565	0.224	77.13	0.717	0.298	20.48	0.022	0.009	42	4.597	1.883	150
Norfolk	28	1.819	0.723	31.77	1.690	0.783	24.31	0.363	0.123	8.5	8.410	3.890	143
Barnstable	15	0.512	0.380	127.73	0.246	0.159	27.29	0.099	0.121	89	1.023	0.685	180
Nantucket	1	0.199	0.193	112									

Table 6: Population (Pop) and housing (Hous) densities per square mile, Source: US Bureau Census (2000), Calculus: authors; Shortest distance to Boston in km (Dist) Source:www.viamichelin.com.

References

[1] Albert, J. (2009): “*Bayesian Computation with R, 2 ed.*” Springer-Verlag, New York.

[2] Alperovich, G. and J. Deutsch (1994): “Joint Estimation of Population Functions and the Location of the Central Business Districts”, *Journal of Urban Economics,* vol. 36, pp. 239-248.

[3] Casella, G. and George, E.I. (1992): ”Explaining the Gibbs sampler”, *American Statistician,* vol. 46, pp. 167-174.

[4] Chotikapanich, D., D.S. Prasada Rao, K.K. Tang (2007): ”Estimating income inequality in China using grouped data and the generalized beta distribution”, *Review of Income and Wealth,* vol. 53(1), pp. 127-147.

[5] Fujita, M. and Thisse, J.F. (2002): ”*Economics of Agglomeration. Cities, Industrial Location and Regional Growth”,* Cambridge University Press, Cambridge.

[6] Gelman, A. and Rubin, D.B. (1992): ”Inference from Iterative Simulation Using Multiple Sequences”, *Statistical Science,* vol 7, pp. 457-472.

[7] Glaeser, E. L. (2008): ”*Cities, Agglomeration and Spatial Equilibrium”,* Oxford University Press.

[8] Gupta, Arjun K. and Nadarajah, Saralees (2004) ”*Handbook of Beta Distribution and Its Applications”,* Dekker, New York.

[9] Hougaard, P. (2000) ”*Analysis of multivariate survival data”,* Springer-Verlag, New York.

[10] Lehmann, E.L. (1966): “Some concepts of dependence”, *The Annals of Mathematical Statistics,* vol 37, pp. 1137-1153.

[11] Mas-Colell, A., M.D. Whinston and J.R. Green (1995): ”*Microeconomic Theory”,* Oxford University Press.

[12] McCullagh, P. and Nelder, John A. (1989): “*Generalized Linear Models”,* Chapman and Hall, London.

[13] McDonald, J.B., Y.X. Yexiao (1995): ”A generalization of the beta distribution with applications”, *Journal of Econometrics,* vol. 66, pp. 133-152.

[14] Nairn, A. G. M. and O’Neill, G.J. (1988): ”Population Density Functions: A differential Equation Approach”, *Journal of Regional Science,* vol. 28(1), pp. 89-102.

[15] Nelsen, R.N. (1999): *"An Introduction to Copulas",* Lecture Notes in Statistics, vol. 139, Springer-Verlag, New York.

[16] Plummer, M. (2009): JAGS Version 1.0.3: Just Another Gibbs Sampler. http://www-fis.iarc.fr/ martyn/software/jags/

[17] R Development Core Team (2009): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[18] Ramcharan, R. (2009): "Why an economic core: domestic transport costs", *Journal of Economic Geography,* vol. 9(4), pp. 559-581.

[19] Sahu, S. K., Dey, D. K., Aslanidou, H. and Sinha, D. (1997) "A Weibull regression model with gamma frailties for multivariate survival data", *Lifetime Data Analysis,* vol. 3, pp. 123-137.

[20] Song, Sh. (1996): "Some Tests of Alternative Accessibility Measures: A Population Density Approach", *Land Economics,* vol. 72(4), pp. 474-482.

[21] World Bank (The) (2008): *"World Developing Report. Reshaping Economic Geography",* The World Bank, Washington D.C.