

Modelling population density over time: How spatial distance matters

Ilenia Epifani* and Rosella Nicolini^{†‡}

August 17, 2015

Abstract

This study provides an empirical application of the Bayesian approach for modelling the evolution of population density distribution across time. It focuses on the case of Massachusetts by tracking changes in the importance of spatial distance from Boston concerning citizens' choices of residence according to data for 1880-1890 and 1930-2010. By adopting a Bayesian strategy, results show that Boston reinforced its attractiveness until the 1960s, when the city's accessibility no longer represented the unique determinant of population density distribution. Referring to selected historical evidence, a few possible interpretations are presented to endorse these results.

Keywords: Bayesian inference, Distance, Ethnic composition, Population distribution, Random frailties.

JEL Codes: C11, N92, R19; *AMS Classification:* 62F15, 62P20.

*Politecnico di Milano, Dip. di Matematica, P.zza L. da Vinci, 32, I-20133 Milan (Italy)

[†]Corresponding author: Departament d'Economia Aplicada, Universitat Autònoma de Barcelona, Edifici B – Campus UAB, 08193 Bellaterra (Spain). E-mail: rosella.nicolini@uab.eu.

[‡]We are grateful to three anonymous referees, D. Cuberes, J. Deveraux, M. Felix, A. Saiz and the participants to the NARSC conference (Atlanta, 2013) and seminar at Universidad Autònoma de Madrid for useful comments and suggestions. We thank E. Cittadini and J. Gaviglio for their technical assistance with the data treatment. All remaining errors are our own responsibility. Financial support from ECO2014-52506-R, 2014SGR327 and XREPP is gratefully acknowledged.

1 Introduction

Individual preferences for residential location rely on a combination of a few determinants ranging from individual-specific characteristics to predetermined neighbourhood characteristics and including subjective beliefs about the behaviour of other individuals (DURLAUF, 2004).

In the literature (e.g., PARR, 1985a; PARR 1985b), when individuals prefer one location over other options, their distribution of choices across space is not uniform; for instance, population density is not identical in urban and rural areas. The distribution of population density is empirically measured as the combination between the value of population density at a given location and its distance from a selected point usually identified as the central business district (CBD), which usually includes important sites for people’s professional and leisure activities. As such, the concept of accessibility to a CBD turns out to be a key feature for modelling population density distribution. In QUIGLEY (1985), accessibility is identified by the presence of infrastructural elements that guarantee individual mobility. In the case of an efficient transportation infrastructure, people may be less interested in being located near the CBD, since proximity to the CBD often entails the presence of congestion effects that, if significant, can urge people to consider settling slightly farther away. On this point, local authorities consider that investing in the construction of effective transportation infrastructure can encourage people’s relocation across space and, in turn, reduce the impact of congestion-related costs.

In this sense, people’s propensity to settle close to a CBD can be read as their desire to be near the focal point of their activities. By contrast, their propensity to move away from it can indirectly indicate the ease of reaching the CBD in terms of transportation costs.

This study focuses on the evaluation over time of the importance of accessibility to Boston in shaping Massachusetts’s population density distribution once Boston has been identified as the CBD. Instead of focusing on changes affecting Boston’s urban structure, this study more accurately concerns how Boston as a pole of attraction has shaped population density distribution in Massachusetts.¹

¹Boston could be also a pole of attraction for other neighbouring areas close to Massachusetts’s borders. However, this study’s analysis privileges the historical dimension and avoids any change at the institutional level (e.g., governance issues or differences in civil legislation across US states) that might distort the identification strategy. Focusing only on Massachusetts, it is controlled for, while the Bayesian approach also deliver insights that can be of interest for Massachusetts’s neighbouring areas.

In this research, the spirit of traditional analysis of population density distribution for the specific case of Massachusetts is obtained, yet it also introduces two novelties with respect to current literature. First, following EPIFANI and NICOLINI (2013), the population density function is considered to be a random variable. Secondly, the analysis is performed by incorporating the time dimension.

As discussed in NAIRN and O'NEIL (1988), adopting a probabilistic density function is a generalisation of the idea of population dispersion with respect to the distance from a centre of attraction (e.g., the CBD). Economic literature (QUIGLEY, 1985; DURLAUF, 2004; TOPA and ZENOU, 2015) identifies the physical distance from a selected site as a determinant for location choice, though such is not the only one. For instance, natural amenities, the quality of public services, and/or the ethnic composition of the neighbourhood can matter as well. As such, working with a probabilistic distance function is a flexible device for combining these types of subjective priority.

Previously, EPIFANI and NICOLINI (2013) exploited a similar method of analysis to assess Boston's centrality in the location preferences of Bay Staters and to identify determinants of population density distribution in a probabilistic setting. By focusing on only the year 2000, these authors identified the physical distance from Boston as the relative dominant determinant of population density distribution, followed by the ethnic composition of the territory and a group of other covariates, including education, age composition, and presence of natural amenities. As a result of this outcome, the present study aims to extend the framework of analysis by including the dimension of time.

This study focuses on the importance of the dimension of accessibility -namely, the importance of distance- in shaping the location decisions of citizens across time. To achieve its scope, an original database has been built by exploiting a set of comparable data extracted from different editions of the US Census. Furthermore, in the tradition of the regional economic literature (e.g., PARR 1985b; NAIRN and O'NEILL, 1988), the study takes as a reference a monocentric distribution function to model the distribution of population density in Massachusetts, where Boston is assumed to be a persistently attractive pole across time. This selection is underpinned by previous studies and empirical evidence.

As documented in GLAESER (2005), it is clear that people have desired to live in Boston since its foundation. Boston's comparative advantage lies principally in its human

This issue is discussed in Section 5.

capital, with whom the city has been able to reinvent itself yet maintain its attractiveness despite having experienced cyclical waves.

Boston's attractiveness with respect to the rest of Massachusetts increased until 1920, when it began to decline during the middle of the 20th century, only to rebound in the 1980s. The reasons for the first period of growth are strongly associated with technological advances that prompted the subsequent urbanisation of industrial activity via requirements imposed upon factories to reduce the space they occupied. Meanwhile, the implementation of a quality public transportation network made it possible to travel around Boston more cheaply than around other low-density communities. However, two principal causes of the decline in the city's attractiveness are rooted in the decline of manufacturing activity and in the general improvement of private transportation: the widespread use of automobiles to improve individual mobility, which favoured population displacement to the city's outskirts. Since the 1980s, Boston has regained its attractiveness, for the abundance of skilled labour in the population made the city highly attractive for skill-specialised activities. Consequently, very stringent urban housing regulations inflated the real estate housing bubble. GLAESER and WARD (2009) provided a comprehensive interpretation of the causes and consequences of housing regulations in Boston, in which the decline in the supply of new homes was not a consequence of a lack of land. In fact, the most effective measure was the definition of the minimum lot size, which in turn encouraged greater housing density. The authors showed that from a historical perspective, towns with more immigrants have had less stringent minimum lot-size regulations. In general, the practice of defining high minimum lot sizes was used by white natives to restrict home construction for blacks and foreigners.

To achieve the study's aims, its strategy of analysis exploits a Bayesian frailty estimation technique, which is a flexible framework that accommodates controlling for heterogeneity. Furthermore, the introduction of priors allows controlling for problems of weak data (LeSAGE and PACE, 2009). In particular, in the definition of the estimation strategy, the study relies upon the possibility of elucidating ad hoc priors in order to embed the dimension of time. Moreover, subjective priors are used to model individual expectations and create dependence across time in modelling population density distribution.

The study's sample selection is limited by the need to identify a sample consistently available over time that allows for comparison. Given the length of the study's timespan, a list of 351 municipalities in Massachusetts has been the focus. These municipalities,

mostly urban areas, are spread across all of the state's territory. With this sample, the study tracks population density distribution across Massachusetts at the urban level from 1930 to 2010, plus two other isolated years, 1880 and 1890, which are considered as external benchmarks for discussing the results of estimations for potentially remote years.

Given the study's interest in analysing changes in the importance of accessibility (i.e., physical distance from Boston) in shaping people's location preferences, it is expected that the improved efficiency of the transportation network during the period monotonically reduces the relative impact of physical distance as a location determinant. Instead, the investigation reveals a bell-shaped dynamics. According to the study's estimations, the centripetal force of Boston as the principal location-attractor in Massachusetts was extremely strong at the end of the 19th century and continued to strengthen until the 1960s, after which it progressively declined. In this respect, the study's results align with those of GLAESER (2005). By contrast, ethnic composition acquired greater relative importance as a location determinant. The ethnic population composition -here measured as the proportion of white people in the total local population- does not importantly impact modelling location choices until the 1970s. The estimated magnitude of its coefficient twists around zero, after which it immediately increases. The study therefore investigates the latter result from a geographical and historical perspective. In reference to current literature, the study proposes some feasible interpretations.

The contribution of COLLINS and MARGO (2011) and FETTER (2012) are good references, for they have documented how the Equal Credit Opportunity Act in the 1970s impacted the US housing market, and, in turn, produced important changes in the ethnic composition of US neighbourhoods (cf. GLAESER and WARD, 2009).

Empirical results also permit tracking the variation of some random spatial county effects as a further determinant in modelling uneven population density distribution across space. The study's findings confirm that the populations of towns belonging to remote counties far from Boston consolidated low attractiveness in relation to Boston as the CBD across time.

In what follows, Section 2 introduces the data, Section 3 describes the empirical strategy, and Section 4 reports estimation outcomes, which Section 5 discusses. Later, Section 6 and Section 7 offer a discussion and conclusions, respectively. Details about econometric estimations appear in Appendices A,B,C.

2 Data selection

To conduct this analysis, a database was constructed that allows examining population density distribution and its selected determinants across time. The primary source is US Census data, and the sample consists of the entire population of the 351 municipalities in Massachusetts in the state’s 14 counties. As for the county composition of Massachusetts, a descriptive map is provided in Figure 3 in Appendix B. The term *municipality* is typical of Massachusetts, where municipalities are established by a general act with a minimum requirement of 12,000 inhabitants. At present, Massachusetts’s municipalities include 39 city and 312 town governments, whose identification is based on the existence of local, established institutions to exert similar power and perform similar functions to govern civil life. According to the historical information provided by the United States Census Bureau,² the structure of the census has experienced relatively few changes across time. In fact, changes that do appear are mostly associated with the growth of new towns or the merging of others already in existence. The first six censuses (1790-1870) were managed by the juridical branch of the federal government. Each district was assigned a marshal who hired local assistants, typically residents from a village or neighbourhood, to conduct the census. During this period, racial classification was carried out by the enumerator, and people were not interviewed. Since 1880, the census has been managed by the executive branch of the federal government. Notably, the 1890 census was the first to be organised by a tabulated scheme with precise questions addressing race answered by people interviewed. Until the 1960s, data collection was conducted by an enumerator and included interviews, while since the 1970s a self-enumerator practice has been adopted. For this, a census questionnaire with instructions has been delivered by mail. Once returned, the forms are reviewed by a census enumerator, and if needed, follow-ups were performed. As for the accuracy of the data, US Census descriptives confirm the consistency of the design of the sample and the elaboration of statistics. Over time, some changes in the structure and convergence of questions have been introduced, though all editions are comparable. Concerning territorial composition, the number of municipalities in Massachusetts was variable until 1930; the sample of the 19th-century census reports only 270 municipalities per year against 351 since 1930.

Concerning determinants of population density distribution, this study’s choice is mo-

²“Through the Decades”, http://www.census.gov/history/www/through_the_decades/.

tivated by reference to current literature. DURLAUF (2004) has discussed that individual-level decisions about location are mostly driven by a combination of three categories of determinants: deterministic individual characteristics, predetermined neighbour-specific characteristics, and individual subjective beliefs. HELSLEY and STRANGE (2007) have since refined that statement by arguing that the most relevant determinants in location decisions are distance from the CBD, quality of the residential area, and accessibility of other places of interest. These results are a generalisation of empirical results discussed in QUIGLEY (1985) for the Pittsburgh metropolitan area, for which the author found two main determinants shape individual location choices: the commute to citizens' principal places of interests (e.g., the workplace) and the racial composition of the neighbourhood as the most salient feature of neighbourhood characteristics.

Given these outcomes and accounting for the geographical structure of the sample drawn from urban areas in Massachusetts, the study's source of data allows considering the distance from the common centre of interest (i.e., Boston), the presence of natural amenities, and the racial composition of municipalities as location determinants. In addition, the study includes some random effects to control for unobserved features.

By incorporating the relative ethnic composition of a municipality, some location preferences are implicitly accounted for that may relate to ethnic or network features that drive individual location choices. For instance, a territory highly dense in white population can be more attractive for whites as opposed to blacks. This idea relies on the propensity of individuals to interact or cluster with other individuals of the same ethnicity given their common cultural habits (TOPA and ZENOU, 2015).

By contrast, not all censuses provide complete, detailed quantitative information of the racial composition of the population. On the contrary, the division between whites and blacks is nearly always present in all census editions since the inception of the institution. More detailed information about other ethnic groups has been added in the most recent years, though such plays a minor role in a great portion of municipalities. For the scope of this research, the study privileges a strategy that splits the sample into two groups: whites and other minorities, with the acknowledgment that blacks are predominant among minority groups. To be consistent over time, the percentage of white people over the total population is used as a synthetic measure of a municipality's ethnic composition. QUIGLEY (1985) has stated that this type of division -namely, whites versus other ethnic groups- turns out to be extremely effective in estimations. However,

one unusual issue about the information remains concerning; in the 1950 census, the word colour was removed from racially oriented questions and reintroduced in 1960. Therefore, testing the impact of ethnic composition in shaping population density distribution in 1950 was impossible.

As a proxy for accessibility, the geodesic distance of each municipality from Boston provided by ArcView has been used. Furthermore, a proxy for natural amenities to represent the quality of life in a municipality has been included. Here, amenities are measured as the proportion of water areas over the total territory in each municipality, which has been treated as constant over time. In the case of the proxy for amenities, this working hypothesis is not extremely binding; in Massachusetts, nearly all artificial lakes and ponds (e.g., the Quabbin and Wachusett Reservoirs) were created or completed in the 1930s, which aligns with the data to which the study refers (SIMCOX, 1992).

As discussed in EPIFANI and NICOLINI (2013), all of these variables jointly model citizens' desire for close proximity to Boston. As such, by estimating a type of population density distribution function year by year, the relative importance of every covariate can be assessed and their evolution tracked over time. Furthermore, the way in which priors are modelled can allow a device to include the dimension of time: by creating a mechanism to introduce individual expectations. Finally, some random effects control for poorly identified territorial features and omitted variables.

3 A hierarchical gamma model

Population density distribution is modelled by referring to a monocentric distribution around a CBD: in this case, Boston. Economic literature has identified Boston as the pole of major attraction in Massachusetts since its foundation (e.g., in GLAESER, 2005). Recently, EPIFANI and NICOLINI (2013) have provided quantitative estimations in which Boston also emerges as the most attractive urban area in Massachusetts. The current study's initial working hypothesis is thus that all residents of Massachusetts consider Boston to be the state's centre of interest, and as their point of attraction, they are willing to relocate to be near it. However, identifying residents' preferences in terms of location choices needs to be modeled by accounting simultaneously for other factors identified with proxies for natural amenities and ethnic composition. The empirical strategy

thus involves associating in a probabilistic manner the population density distribution in Massachusetts with a group of selected variables representing factors describing people's location preferences.

In this analysis, a parametric model is implemented. The choice of likelihood for the municipality's population density is based on careful observations of the nature and properties of the data. A preliminary explorative data analysis of population density at the municipal level in the 14 counties of Massachusetts from 1930 to 2010 is shown in Figure 1, which contains the sample means and standard deviations of population densities of each county on a logarithmic scale.

[Figure 1 about here.]

Figure 1 reveals an increase in the standard deviation with a larger mean of population density during each decade. Furthermore, the mean and variance of municipality population densities exhibit a different trend in each county. The former feature of the data -namely, an increasing standard deviation with larger mean of population density- is consistent with a heteroscedastic model. The latter feature suggests that the introduction of some spatial county effects in the model can capture heterogeneity among counties.

Following EPIFANI and NICOLINI (2013), this study uses a hierarchical gamma model. A gamma function has already been proposed as an appropriate tool for modelling population density distribution across regional space by SONG (1996). However, an alternative to gamma could be represented by a lognormal model, since both gamma and lognormal likelihood are very popular in modelling positive variables and characterised by two parameters, as well as assume a standard deviation increasing with the mean. In fact, it is often difficult to know whether one should assume a gamma or lognormal distribution. McCULLAGH and NELDER (1989) have suggested a gamma assumption for working with data on an original scale instead of on a log scale (cf. PARR, 1985a). Such is this study's case because, for instance, if the linear scale is preserved, then the sum of the population densities of the county's municipalities retains the appropriate mathematical definition of the county's population density. In addition, FIRTH (1988) has argued that a gamma model performs slightly better than the lognormal under reciprocal misspecifications. As such, the lack of covariates to include in the model makes the study's setting more sensitive given the misspecification problem when modelling population preferences.

For this reason, a hierarchical gamma model may provide a suitable simple framework

for conducting a study of the determinants of population density distribution in Massachusetts.³

This study’s Bayesian hierarchical gamma model with random county effects specifies:

a) a gamma *likelihood* for population densities describing the relationship among densities, as well as predictors and random county effects so that population density is gamma-distributed with time-varying parameters each time,

b) a gamma distribution for the random county effects,

c) a *prior* distribution for the unknown parameters in parts a) and b).

The discussion on the prior distribution in point c) is incorporated in Section 4 which concerns the Bayesian estimation. Instead, introducing some preliminary definitions is necessary to give a formal statement of parts a) and b) of the model.

Considering the whole territory of Massachusetts, the population density and its determinants can be defined as follows. For the j th municipality within the i th county, let $Y_{ij}^{(t)}$ be its density of population at decade t , $Dist_{ij}$ is its spatial distance from Boston,⁴ and $Mix_{ij}^{(t)}$ its proportion of white people at decade t . At the same time, Z_i signifies the size of the local amenities measured as the proportion of water areas in the i th county territory,⁵ whereas $w_i^{(t)}$ represents the random effect associated with county i at time t , and n_i denotes the number of municipalities in county i , for $i = 1, \dots, 14$. Counties are identified by their number code in the first column of Table 6 in Appendix A and the decades involved in the analysis are $t = (1880, 1890) \& 1930, \dots, 2010$. Lastly, the notation $X \sim \Gamma(a, b)$ means that a random variable X is gamma-distributed with shape a and rate b .⁶

Equation (1) formally defines the part a) of the model, i.e. the conditional distribution

³As it has been pointed out by a referee, more flexibility may be obtained in a nonparametric way by, for instance, a mixture of distributions. Nevertheless, this kind of specification is analytically much more demanding.

⁴In this study we privilege to focus on this measure of distance rather than on others measures as the travel-time, for instance. Our choice is driven by two orders of reasons. First, we do not dispose of available data about the travel-time for a so long period of time. Secondly, we are performing year-by-year estimations. In order to exploit the variability of the travel-time (and hence, the information delivered by this variable) it is necessary to deal with panel-style data.

⁵GLAESER and WARD (2009) argue that water is an important amenity for creating recreational spaces and important laws have been passed to protect waterways and wetlands.

⁶The gamma probability density $\Gamma(a, b)$ with shape a and rate b has kernel: $x^{a-1}e^{-bx}$ on the half-line $(0, \infty)$ and it is equal to zero otherwise. Expected value and variance of a random variable $X \sim \Gamma(a, b)$ are given by $E(X) = a/b$ and $\text{Var}(X) = a/b^2$ so that the coefficient of variation (i.e. the ratio between the standard deviation and the mean) is $\text{Cv}(X) = 1/a$.

of the $Y_{ij}^{(t)}$'s given the random effects $w_i^{(t)}$ and the part b), i.e. the law of $w_i^{(t)}$:

$$\begin{aligned}
 Y_{ij}^{(t)} | w_i^{(t)} &\sim \Gamma(\theta^{(t)}, \theta^{(t)} w_i^{(t)} / \mu_{ij}^{(t)}) \text{ independent } \forall i, j, t \\
 (1) \quad w_i^{(t)} &\sim \Gamma(\alpha^{(t)}, \alpha^{(t)}) \text{ independent } \forall i, t \\
 \mu_{ij}^{(t)} &= e^{\beta_0^{(t)} + \beta_1^{(t)} Z_i + \beta_2^{(t)} Z_i Dist_{ij} + \beta_3^{(t)} Dist_{ij} + \beta_4^{(t)} Mix_{ij}^{(t)} + \beta_5^{(t)} Mix_{ij}^{(t)} Dist_{ij}}
 \end{aligned}$$

The predictor $Dist_{ij}$ captures the impact of the distance from Boston as a factor driving the distribution of population density. According to the hypothesis underpinning Equation (1), $Dist_{ij}$ embeds the attractiveness of the capital and, moving away from Boston, lower values of population density should be noted. By contrast, Z_i controls for the geographic characteristics of a territory i while $Mix_{ij}^{(t)}$ controls for the ethnic composition at municipality level. It should be noted that estimations are run separately by year and therefore use cross-sectional data. As such, municipality-level fixed effects cannot be incorporated to capture municipality-level heterogeneity. The $Mix_{ij}^{(t)}$ covariate plays an important role in partially controlling heterogeneity across municipalities and embeds the attractiveness of the municipality for a specific ethnic group (namely whites) which can derive from different reasons, including specific policies implemented at local level (e.g. zoning laws). The sign of this variable is expected to be negative when the white population aims at settling far from the CBD to live in individual dwellings and for instance enjoy good accessibility for reaching the CBD.

The econometric Equation (1) also includes an interaction term $Dist \times Z$ between the distance from Boston and the natural amenities as well as an interaction $Dist \times Mix^{(t)}$ between the distance from Boston and the ethnic composition. While the former term: $Dist \times Z$ aims to control for the trade-off between the preference for high concentration of natural amenities -usually enjoyed far from Boston- and the desire to be near Boston, the latter: $Dist \times Mix^{(t)}$ controls for the interplay between the ethnic composition of a municipality and, again, the distance from Boston. As such, the extent to which a measure of the ethnic composition of a territory -here, the density of whites at the municipality level- is more or less exacerbated in location decisions by the distance from Boston. The interaction term $Dist \times Mix^{(t)}$ is expected to impact population location decisions when the quality of the infrastructure allows for comfortable commutes and makes the variable $Dist$ (i.e. transport costs) less dominant. To assess the importance of $Dist$ and $Mix^{(t)}$, it is first assumed that $\beta_5^{(t)} = 0$, and later $\beta_5^{(t)} \neq 0$.

Finally, an unobserved frailty $w_i^{(t)}$ for each county i and decade t is also incorporated in Equation (1) to capture *i*) the degree of similarity of population habits (or similar regulation in civil matters) for people living in municipalities belonging to the same county as well as *ii*) the degree of heterogeneity among the counties. In one way or another, each county frailty $w_i^{(t)}$ summarises all the predictors of the population density, both unobservable and observable, but not explicitly taken or accounted for by the remaining covariates in the i th county at time t . As such, the county frailties can be assimilated to a sort of random spatial effects. Similar to EPIFANI and NICOLINI (2013), this analysis assumes stochastic independency between frailties $w_1^{(t)}, \dots, w_{14}^{(t)}$, given the independence of each county land organisational structure. At the same time, there is dependency in each county, meaning that municipalities belonging to the same county share some common features that in our model are associated with the fixed effects embedded in Z 's and the random effects $w_1^{(t)}, \dots, w_{14}^{(t)}$.

Since the conditional expectation of $Y_{ij}^{(t)}$ given the county effect $w_i^{(t)}$ is $E(Y_{ij}^{(t)}|w_i^{(t)}) = \mu_{ij}^{(t)}/w_i^{(t)}$, then a gamma frailty $w_i^{(t)}$ significantly smaller than one amplifies the impact of the predictors on the population density distribution in county i at time t , whereas a gamma frailty $w_i^{(t)}$ "far greater than one" collapses them and, in a very extreme situation, $w_i^{(t)} \approx 1$ excludes the existence of territorial differentiation.

The values of $w_1^{(t)}, \dots, w_{14}^{(t)}$ are probabilistically controlled by a single time-variant parameter $\alpha^{(t)}$ in such a way that the mean of every $w_i^{(t)}$ equals unity while its variance is $1/\alpha^{(t)}$. This dynamics implies that for large values of $\alpha^{(t)}$ all $w_1^{(t)}, \dots, w_{14}^{(t)}$ are a priori concentrate around one, resulting in a condition of lack of territorial differentiation, save some possible morphological differences between county areas recorded in the natural amenities Z 's. Instead, small values of $\alpha^{(t)}$ cause a skewed distribution of the $w_i^{(t)}$'s that permits for them both large and small values, and then represent a strong heterogeneity among counties at time t . Not least, $\alpha^{(t)}$ is dimensionless and has the direct meaning of variability measured by the coefficient of variation. This probabilistic modelisation of the frailties is similar to the prior distribution used, for example, by GEWEKE (1993) to model heteroscedasticity (and outliers) in Bayesian linear regression and by LeSAGE and PACE (2009) in the context of Bayesian heteroscedastic spatial models. According to the latter model, all observations are split into n groups according to n points in space where they have been collected and their variances may vary across space. The prior distribution for the unknown variances takes the form of a set of n iid inverse chi-squared

$\chi^2(r)/r$ distributions, in which r represents the single parameter of the $\chi^2(r)$ distribution.

The parameters $\alpha^{(t)}$ along with $\theta^{(t)}$ also measure the intensity of the relationship among population densities of different municipalities within the same county. It can be proven that the linear correlation coefficient between two of them is given by

$$\rho(Y_{ij}^{(t)}, Y_{ih}^{(t)}) = \frac{\theta^{(t)}}{\theta^{(t)} + \alpha^{(t)} - 1}, \quad \text{for all } t, i \text{ and } j \neq h$$

as shown in EPIFANI and NICOLINI (2013). In particular, for large values of $\alpha^{(t)}$, the correlation of the municipality population densities within the same county approaches zero. By contrast, small values of $\alpha^{(t)}$ represent a strong positive relationship of densities among municipalities in the same county. Furthermore, for given values of $\alpha^{(t)}$, the larger $\theta^{(t)}$ is, the larger the dependence between the population densities of the municipalities within the same county.

The parameters $\alpha^{(t)}$ and $\theta^{(t)}$ also control the variability of the density population. Indeed

$$\text{Var}(Y_{ij}^{(t)} | w_i) = \frac{1}{\theta^{(t)}} \left(\frac{\mu_{ij}}{w_i} \right)^2$$

and

$$\text{Var}(Y_{ij}^{(t)}) = \text{E}(\text{Var}(Y_{ij}^{(t)} | w_i)) + \text{Var}(\text{E}(Y_{ij}^{(t)} | w_i)) = \frac{\alpha^{(t)} - 1 + \theta^{(t)}}{\theta^{(t)}(\alpha^{(t)} - 2)} \text{E}(Y_{ij}^{(t)}), \quad \text{for } \alpha^{(t)} > 2$$

As such, the larger $\alpha^{(t)}$ and $\theta^{(t)}$ are, the smaller the “marginal” variances $\text{Var}(Y_{ij}^{(t)})$ of the population densities and then the more concentrated the $Y_{ij}^{(t)}$.

As concerns regression parameters, coefficients $\beta_3^{(t)}$ associated with the distance are expected to be negative for all t , as it is commonplace in the literature. Similarly, the $\beta_4^{(t)}$ associated with $Mix^{(t)}$ will be negative whenever the ethnic composition of a territory is a true discriminating factor in location choices (as previously argued, as in QUIGLEY, 1985).

The unknown parameters of the Equation (1) are $\beta_0^{(t)}, \beta_1^{(t)}, \dots, \beta_5^{(t)}, \theta^{(t)}, \alpha^{(t)}$ along with the unknown county frailties $w_1^{(t)}, \dots, w_{14}^{(t)}$ for $t = 1930, \dots, 2010$. Estimations are performed in three steps beginning by running the regression including only distance and frailties for each year in the sample and assuming that $\beta_1^{(t)} = \beta_2^{(t)} = \beta_4^{(t)} = \beta_5^{(t)} = 0$ in Formula (1). The purpose of this first exercise is to track the evolution of the coefficients related to the distance from Boston as the unique determinant of population density dis-

tribution. This model is labelled *Baseline model (BM)*. Thus, considering an alternative reduced version of Equation (1), without the interaction $Dist \times Mix^{(t)}$ (namely $\beta_5^{(t)} = 0$) a model is produced, labelled *Model with Ethnic Composition (MEC)*. Ultimately, estimations for the complete model described in (1) -from now on called *Model with full interactions (MFI)*- are run. In this way, the significance and variation of the coefficients associated with the two expected principal covariates across time (i.e. distance from Boston and ethnic composition of the population) are first tracked by progressively including more controls. Secondly, the evolutionary trend of the coefficient of distance in the three alternative reductions of the hierarchical Model in (1) are compared to capture variation of the magnitude of coefficients of some selected covariates across time. In this sense, the analysis seeks to assess the magnitude of changes for the estimated coefficients associated with accessibility and neighbourhood characteristics that reflect changes in their relative importance in influencing location choices.

All the models were run from 1930 to 2010, but the lack of data on the ethnic composition restricted our estimations for 1950 to *BM* only. To formulate some suggestions about the importance of the distance in the location choices at the beginning of census data availability, *BM* was also run for the years 1880 and 1890. In this respect the statistics obtained referring to these two years are not fully comparable to those from 1930 onward. Nevertheless, estimations of the most remote years may partly indicate how dynamics of the distribution of population density over time can be interpreted.

4 Bayesian estimation

This study's empirical analysis occurs within a Bayesian framework. Two motivations especially drive this choice. First, the Bayesian paradigm -in which all that is unknown is probabilistically modelled as randomness- provides a more natural framework for analysing hierarchical models with rich parametrisation and random spatial effects. In the language of the Bayesian inference, the county frailties $w_1^{(t)}, \dots, w_{14}^{(t)}$ are simply unknown parameters with an assigned prior probabilistic law as presented in the second line of Equation (1) that takes the form of a set of 14 independent $\Gamma(\alpha^{(t)}, \alpha^{(t)})$ distributions, for each t . Hence, we can handle them as other unknown parameters of the model, $\beta_0^{(t)}, \beta_1^{(t)}, \dots, \beta_5^{(t)}, \theta^{(t)}, \alpha^{(t)}$. In this way, controlling for heterogeneity between counties along with the term Z_i becomes possible, which is of crucial concern, especially in repli-

cating the same framework of analysis for different moments in time.

Second, a limited number of data (351 for every time t) with respect to the number of parameters to be estimated (8 parameters and 14 county effects for every t) is disposed. However, the Bayesian approach, which combines information from priors and the sample, still works; in the context of hierarchical models, a Bayesian statistician can even estimate more parameters than observations, though such is invalid in frequentist inference.⁷

According to a Bayesian paradigm, the unknown parameters $\beta_0^{(t)}, \beta_1^{(t)}, \dots, \beta_5^{(t)}, \theta^{(t)}, w_1^{(t)}, \dots, w_{14}^{(t)}$, for $t = 1930, \dots, 2010$, are random variables from a *prior* joint distribution $\pi(\cdot)$ which we have to specify, and the statistical problem lies in updating $\pi(\cdot)$ by computing the *posterior* joint conditional probability $\pi(\cdot|\mathbf{Data})$ of the parameters, given the collection of $\mathbf{Data} = (\mathbf{Y}, \mathbf{Dist}, \mathbf{Mix}, \mathbf{Z}) \equiv \{Y_{ij}^{(t)}, Dist_{ij}, Mix_{ij}^{(t)}, Z_i : j = 1, \dots, n_i, i = 1, \dots, 14 \text{ and } t = 1930, 1940, \dots, 2010\}$. The posterior distribution is obtained via the Bayes'rule as

$$\pi(\mathbf{parameters} | \mathbf{Data}) \propto \pi(\mathbf{parameters}) \times L(\mathbf{Data} | \mathbf{parameters})$$

where L is the likelihood function. Then the posterior joint distribution is summarized in a simple way -typically by posterior means- giving rise to point estimates of the unknown parameters. Moreover, the precision of the estimation of the unknown parameters is summarized by means of the associated posterior standard errors and some Bayesian credible intervals.⁸

Neither the joint nor the marginal posterior distributions of the unknown parameters usually have a closed form, as they are required to be solved some integrals that do not admit any analytical solution. However, Markov Chain Monte Carlo (MCMC) algorithms can be used to simulate them. We coded a GIBBS sampling algorithm in the JAGS (Just Another Gibbs Sampler) software package by PLUMMER (2014), which is designed to work closely with the R package. All statistical computations and graphics were performed

⁷See, for instance, <http://www.bayesian-inference.com/samplesize>. Certainly, a Bayesian hierarchical model overcomes the difficulty of dealing with the problem of over-parametrization, but it requires to be cautious in the definition of the priors in order to make them truly representative and also preserve the sensitivity of the model to capture changes due to the variations of the parameters over time. Furthermore, the authors are conscious that a small sample size makes difficult to single out a possible discrepancy between the prior and sample information because the empirical information is diluted on a too large parametric space.

⁸In Bayesian statistics, a γ 100% *credible interval* for an unknown quantity η is given by $q_{(1-\gamma)/2} < \eta < q_{(1+\gamma)/2}$ in which $q_{(1-\gamma)/2}$, $q_{(1+\gamma)/2}$ are posterior quantiles of η .

in the R package.

Prior specifications for $\beta_i^{(t)}$'s, $\alpha^{(t)}$, $\theta^{(t)}$. In order to embed the time dimension, a novelty in the manner to deal with the priors of $\beta_i^{(t)}$'s, $\alpha^{(t)}$, $\theta^{(t)}$ is introduced by adopting two different strategies that lead to diffuse and historical strategies. The former strategy was carried out by assuming that all the regression parameters $\beta_0^{(t)}, \beta_1^{(t)}, \dots, \beta_5^{(t)}$ are a priori independent normal random variables with zero mean and large variance 10 000: $\beta_0^{(t)}, \beta_1^{(t)}, \dots, \beta_5^{(t)} \stackrel{iid}{\sim} N(0, 10\ 000)$ for $t = 1930, 1940, \dots, 2010$. Instead, all the shapes $\alpha^{(t)}$ and the rates $\theta^{(t)}$, for $t = 1930, 1940, \dots, 2010$, are a priori independent exponential random variables with a rate equal to 0.2. Moreover all $\alpha^{(t)}$, $\theta^{(t)}$ are independent of each $\beta_i^{(t)}$. Such choice for $\alpha^{(t)}$ and $\theta^{(t)}$ make them concentrated around 5, but with a large variance equal to 25. Because of the large variance, these priors are as vague as possible, so that the MCMC algorithm widely explores the parametric space, yet still allows the MCMC algorithm to converge well with this study's data and model. This strategy is hereafter referred to as the *diffuse* prior.

By contrast, the latter strategy produces a light time-dependence between parameter estimates. To explain, the independence of the parameters and the form of the diffuse prior densities are both preserved, yet now at each decade t , whereas, the values of the prior hyperparameters are taken according to the posterior estimates given the previous decade. For example, to estimate the parameters with data from the 1940 census, the regression parameters $\beta_0^{(1940)}, \beta_1^{(1940)}, \dots, \beta_5^{(1940)}$ are assumed a priori normally-distributed with means and variances equal to the posterior means and variances of $\beta_0^{(1930)}, \beta_1^{(1930)}, \dots, \beta_5^{(1930)}$, whereas $\alpha^{(1940)}, \theta^{(1940)}$ are a priori independent exponential with means equal to the posterior means of $\alpha^{(1930)}, \theta^{(1930)}$, respectively.

We refer to this second strategy as the *historical* prior. Put differently, the historical prior is quite comparable to the canonical way to deal with the classical idea of adaptive expectations.⁹

Though *BM* and *MEC* have been run under both diffuse and historical priors, the *Model with full interactions (MFI)* has been estimated using only the diffuse one. At any rate, looking at the results we obtained under diffuse priors, we can appreciate that

⁹This historical strategy includes only marginal information derived from the posterior distribution given the data of the previous decade. Including information from the past about the dependence structure in the prior distribution when non normal-distributed parameters are involved is a technical issue that is difficult to resolve. We leave it for further investigation.

MFI confirms the outcomes obtained in *MEC* as for our parameters of interest. Given this evidence, the estimations of *MFI* can be thought as a sort of exercise of robustness check.

Comparing the estimation outcomes from the two alternative strategies allows testing the potential effect of unexpected variations in the conditions underlying location choices. In the case that the two estimates obtained under the two kinds of priors turn out to be almost identical, such similarly would mean that no important events affected citizens' belief conditions for location choices. If the two types of estimates differ, then the difference would imply that citizens adopted a sort of adaptive behaviour in their decision-making process as a consequence of an important event that affected their decision making from a temporal perspective.

5 Estimation results

To manage tractable values, the predictors *Dist*, $Mix^{(t)}$ and *Z* were standardized by subtracting their sample mean and dividing by their sample standard deviation.

On the whole, 750,000 iterations for three chains were run for estimating the unknown parameters in the three models *BM* or *MEC* or *MFI* for each decade *t*, and the first 250,000 were discarded as burn-in. After the burn-in, one out of every 100 simulated values was kept for posterior analysis, for a total of 5,000 simulations saved per chain for each time *t*. A final sample was selected among these three. The convergence diagnostics, such as those available in the R package CODA (Gelman, Geweke, Heidelberger and Welch stationarity test, interval halfwidth test) were computed for all parameters, indicating that convergence was achieved.

The first concern has been to test the goodness of fit of our empirical specifications. The best matching of the distribution of population density was selected among *BM*, *MEC* and *MFI* according to the Bayesian Deviance Information Criterion (DIC),¹⁰ and the percentage of “Bayesian outliers”:¹¹ the model yielding the smallest value of DIC

¹⁰The DIC is a generalization of Akaike's Information Criterion (AIC). It is given by the deviance (i.e., minus two times the likelihood) calculated in the posterior means of all parameters plus two times the effective numbers of parameters pD . Models with a small DIC are preferred over those with a large DIC.

¹¹Town *j* in county *i* has been classified as an *outlier* at time *t* if its real population density $y_{ij}^{(t)}$ does not fall within the 95 percent posterior credible interval of the population density $Y_{ij}^{(t)}$ given **Data** = *Y, Dist, Mix, Z*. Cf. Appendix C and Figures 6,7 therein for more details on the Bayesian outliers and the 95 percent posterior credible intervals of the population densities.

and/or smallest value of the percentage of outliers was chosen. Table 1 shows that from 1930 to 1970 the simpler *BM* is sufficient for explaining population density. Note that the percentages of outliers for *BM*, *MEC* and *MFI* remain quite small (always under 5 percent) across time. By contrast, according to the DIC criterion,

[Table 1 about here.]

MEC clearly dominates *BM* from 1980 onward, while *MFI* dominates the two others from 1990 onward.

Referring to the literature, these results confirm that Boston was able to regenerate because of its capacity to attract people in the last decades and that the ethnic dimension turns out to be a key factor in explaining the distribution of population density (GLAESER, 2005).

As far as the estimates, Tables 2(a), 2(b) and 3(a), 3(b) present some posterior summaries of $\beta_0^{(t)}, \beta_1^{(t)}, \dots, \beta_5^{(t)}, \alpha^{(t)}$ and $\theta^{(t)}$, for *BM* and *MEC*, under the diffuse and historical priors, respectively.

[Table 2 about here.]

[Table 3 about here.]

Meanwhile, Table 4 reports the results of the estimations for *MFI* under the diffuse prior.

[Table 4 about here.]

The temporal evolution of the regression coefficients is depicted in Figure 2(a) in the case of diffuse priors and in Figure 2(b) in the case of historical priors.

[Figure 2 about here.]

Lastly, the behavior of the posterior means of the random county gamma frailties $w_1^{(t)}, \dots, w_{14}^{(t)}$, under the diffuse prior across time, are shown in Table 5(a) for *BM*, in Table 5(b) for *MEC* and in Table 5(c) for *MFI*.

[Table 5 about here.]

From Tables 2-4 it can first be deduced that, when statistically significant, all estimated regression coefficients display the expected sign. These estimations have been run by including random frailties that allow accounting for omitted variables that cannot be dealt with explicitly.

Results show that the distance from Boston is estimated to be negatively associated with the size of population density, as well as the *Mix* variable. Proximity to Boston and ethnic composition at the municipality level therefore matter in residential location decisions in our sample. In general, white people seem to appreciate locating in less dense areas and to show a prominent preference for living in individual dwelling properties. The presence of natural amenities does not play any role in residential location decisions in the case of diffuse priors, though under historical priors, it is estimated to have become significant and with a negative sign in the most recent decades, meaning that there is not really an interesting asset for location choices. At the same time, the interaction term $Dist \times Z$ appears to be significant only in the case of historical priors, while $Dist \times Mix^{(t)}$ is statistically significant only for the most recent decades, namely since 1990 (Table 4). When statistically significant, the negative sign of the interaction $Dist \times Mix^{(t)}$ reinforces the idea of the importance of the ethnic composition of the municipality in residential location choices and thus in evening out the relevance of the distance from Boston.

Concerning the temporal evolution of the distance effect, Figure 2(a) and Table 2(a) reveal that in the case of diffuse priors, the trend of the distance coefficient $\beta_3^{(t)}$ is not monotonic over time. The magnitude of $\beta_3^{(t)}$ records varies up to 50% throughout the period. The centripetal forces of Boston as Massachusetts's CBD were extremely strong at the end of 19th century, and their reinforcement continued up to the 1960s. However, the coefficient of the population composition index (i.e., the *Mix* covariate) did not have an important impact before the 1970s; its magnitude twists around 0 as a statistically significant coefficient and becomes negative with a monotonically decreasing slope from 1970 onward. In the 1970s, Boston was no longer a strong attractive CBD for white people, who then preferred to settle far from it, though it was attractive to them until the 1950s. The evolution of the significance of the distance in *MEC* is especially interesting. The trend of the coefficient of the covariate distance in *MEC* follows an inverted U-shaped curve from 1930 onward. Until 1970, there is no remarkable difference between the estimates of the distance coefficients in *BM* and *MEC*. Yet, if the two estimations of the distance effect under diffuse priors Figure 2(a) and under historical priors in Fig-

ure 2(b) are compared, it is clear that the trend is more or less identical. More precisely, it is smoother in the case of historical priors than in the case of diffuse ones,¹² and the magnitude of the coefficient is larger in the case of historical priors than in the case of diffuse ones.

However, as soon as the *Mix* covariate becomes significant, the two estimates of $\beta_3^{(t)}$ under *BM* and under *MEC* diverge. In the most recent years, the value of the distance effect in *BM* is basically constant, though it decreases impressively in *MEC*. This result suggests an interesting interpretation; there is an important expectation component associated with the spatial distribution of ethnic groups that impacts residential location choices in the most recent decades when this effect is more relevant. It is also worth noting that in the case of historical priors, the estimates of amenities Z and the interaction term between distance and amenities are both positive and statistically significant (Table 3(b)).

If spatial dimension were a key factor, the estimation of the county frailties should reinforce the interpretation of the previous results. The frailty terms play an important role, for they correct the magnitude of the decrease in population density while moving farther from Boston. In particular, the posterior means of the random county frailties $w_1^{(t)}, \dots, w_{14}^{(t)}$ shown in Tables 5(a)-5(c) are computed under the diffuse prior. Instead, Figures 4 and 5 in Appendix B depict 90 percent credible intervals for the gamma county frailties $w_1^{(t)}, \dots, w_{14}^{(t)}$, under the diffuse and historical priors, respectively. These figures can be used for a robustness analysis with respect to the diffuse or historical priors, as well as for an analysis of the precision of estimates of $w_1^{(t)}, \dots, w_{14}^{(t)}$. Colored map of Massachusetts in Appendix B, Figure 3, shows that counties with frailties significantly less than 1 are far from Boston, except Suffolk. On the one hand, this outcome emphasises that population in municipalities belonging to counties far from Boston might be attracted by CBDs other than Boston and located in neighbouring counties or states.¹³ On the other hand, the case of Suffolk, the county in which Boston is located, might be explained by the potential existence of another considerable point of attraction close to Boston. Actually, in EPIFANI and NICOLINI (2013), DIC exercises to identify Boston as Massachusetts's CBD revealed the presence of other towns still in Suffolk or neighbouring

¹²From a statistical viewpoint, this circumstance stems from the low variance of all historical priors, given their dependence on the efficient and concentrated estimates of the previous period.

¹³Further investigation in this direction implies moving from a monocentric distribution function to a polycentric one. Nevertheless, given that the most geographical remote areas in Massachusetts are less sensitive to Boston's attractiveness is a clear signal that the geographical dimension considered here (i.e., Massachusetts) fits the analysis. In this sense, border states do not de facto limit the value of the results.

counties (e.g., Quincy in Norfolk County) that act as local points of attraction and whose importance approaches Boston’s.

In addition, counties with a frailty term far greater than 1 are those near Suffolk and thus reflect Boston’s attractiveness. However, for both *BM* and *MEC*, the random county effects seem to vanish over time, since their values concentrate close to 1 in the most recent decades. In parallel, a progressive homogenisation effect seems to occur across time for these countries, which makes the remaining covariates of estimations increasingly determinant in shaping population density distribution. The only exception is represented by Barnstable County, which exhibits an opposite trend, for there was no clear Barnstable effect until 1970. Instead, since 1980 the impact of distance from Boston and *Mix* predictors on population density are amplified by a county random frailty $w_{13}^{(t)}$ concentrated around 0.65, which signifies a peculiar county effect in this county’s municipalities from 1970 onward.

Furthermore, as shown in Tables 5(a), 5(b) and 5(c), for nearly all counties, the county random effects are less important in *MEC* and *MFI* than in *BM*. This evidence is especially remarkable from 1980 onward, when *MEC* and *MFI* are actually preferred to *BM*, which could signify the increasing importance of the ethnic predictor $Mix^{(t)}$ in describing population density distribution across time.

Focusing on $\alpha^{(t)}$, it is observed that under the diffuse prior the $\alpha^{(t)}$ are all concentrated around 5 (with a large variance equal to 25). Instead, the posterior estimates of $\alpha^{(t)}$ in Tables 2 and 3 show that *a posteriori* they become reduced and far smaller posterior values are obtained around 1970, under both *BM* and *MEC*. This result means that until 1970, under both diffuse and historical priors and for both *BM* and *MEC*, the dependence among observations via the county effect reinforces the heterogeneity among counties as well. By contrast, after 1980 the posterior means of $\alpha^{(t)}$ become strongly dependent on the specification of both priors and the model.¹⁴

A reasoned composition of the different pieces of these estimations paints an interesting picture. Since we are working with a monocentric distribution setting, it is important to understand how the changes affecting the centre of the distribution—here, changes in Boston at the economic, social, and/or urban levels—impact the shape of the distribution

¹⁴Given the computational technique, in the case of historical priors, the estimates for $\alpha^{(t)}$ should not be considered fully reliable. The algorithm performs poorly in updating the posterior values of $\alpha^{(t)}$, probably since these posteriors become increasingly more concentrated around a few values from 1940 onwards and then the algorithm may get stuck in the initial values.

(i.e., spatial population distribution across Massachusetts). For one, the attractiveness of Boston as a CBD and Massachusetts's consequent population density distribution are not driven simply by the ease of access to Boston premises. As for the quality of the transportation infrastructure, the Boston area and Massachusetts in general are often considered by analysts as territories with substantial congestion problems. According to data published by IHS Global Inside (2012), congestion costs per auto commuter in the Boston area have greatly increased since 1990. In 2010 dollars, these costs were estimated at roughly 465 in 1990 (in the middle of the US rankings), rose to 938 in 2000, and since hit 980 in 2010, making the city the 9th most expensive in the United States in terms of congestion costs. Plus, the quality of the transportation infrastructure is relatively poor, despite notable investment over time; in 2006, estimates of state-level expenditure on public investment (as a percentage of GDP) assessed the state value at 1.6 percent ranking Massachusetts 45th in the US (HEINTZ et al., 2009). According to HEINTZ et al. (2009), Massachusetts also ranked 8th, 20th, 5th, and 1st in the United States on this count in the periods 1966-1975, 1976-1985, 1986-1995, and 1996-2006, respectively.

In light of this evidence, it is possible to consider that in the last decades, the reduced importance of distance from Boston in shaping Massachusetts's population density distribution, as represented by a drop in the estimated coefficient of the spatial distance covariate, is not uniquely driven by improved transportation infrastructure.

Furthermore, people generally appreciate the existing transportation facilities yet they are greatly concerned about other factors that qualify the quality of life in the neighbourhood in which they live (e.g., DURLAUF, 2004). Economically oriented literature covers this issue quite extensively, and a major explanation for the phenomenon (i.e., the real estate market) may be that property in Boston and its surroundings are extremely expensive, which prompts people to prefer settling elsewhere. GERARDI et al. (2010) present some interesting figures; the secondary market turns out to have been important since the 1980s (40% of new originations) before peaking at 73% in 2005. Until the 1960s, savings were the most importance income source for investment in properties. In the 1960s, inflation and interest rates rose, which drove up the cost of funding for savings and loans. Again, between 1995 and 2005, the Boston area recorded another rise in housing prices, with a 15percent pick-growth in 2000.¹⁵ Additionally, given the difficulty of accessing property then, as a CBD, Boston could have a greater concentration of immigrants who

¹⁵More detailed data can be found at <http://www.forecast-chart.com/>.

may prefer to or have no other option than to rent. This difficulty de facto favours the creation of an ethnic *ghetto* that reinforces the tendency of the white population to settle far from the CBD. Along these lines, BOUSTAN and MARGO (2013) have documented that from 1940 to 1980, black home ownership in US urban centres increased. This tendency was supported by an increase in white home ownership in suburban areas, where the fall in real estate prices improved the owner-occupancy of residential properties.

6 Discussion

This study's estimates show a decline in the importance of the covariate distance from CBD in citizens' residential location choices over time, while the ethnic component has become increasingly important. The interpretation of these results must be associated with an important structural transformation at the territorial level that occurred in the US throughout the period studied here.

Clearly, racial discrimination in the US housing market has been constant over time. CUTLER et al. (1999) document interesting trends:

- During 1890-1940, black people migrated from the rural South to the urban North. In this respect, cities developed with entirely black housing.
- During 1940-1970, the migration of black people expanded across the United States.
- During the 1950s, some white people took action to exclude black people from their residential areas.
- Since 1970, there has been a reduction in racial segregation in residential areas, and black people have moved into once all-white areas.

ROSS and TURNER (2005) detect a persistent discriminatory component in the housing market against African-American and Latino families. These trends induced distortions, possibly to the effect that real estate agencies might have refused to show units in predominantly white neighbourhoods to minorities in order to avoid offending future clients or lowering market activity in key markets. Ethnicity is thus used to signify some unobservable factors of economic returns from market interactions.

As for the mortgage market, there exist several contributions that this study's results support with new, structured quantitative evidence. Referring to the entire United

States, COLLINS and MARGO (2011) focus on the rate of home ownership for black and white households in the US from 1870 to 2007. The narrowing of the racial gap declined by 25 percent from 1870 to 1910, and racial convergence was reinforced until the 1940s. The current study's results align with this evidence. In the presence of a poorly organised transportation system and with manufacturing jobs concentrated in urban areas, workers have preferred to live near their place of employment. Indeed, most have been renters. As the transportation system improved, workers gained mobility. However, the mobility effect must be combined with limitations imposed by the mortgage market. In the 1930s and 1940s, reforms in the mortgage market pushed the consolidation of the practice of redlining black neighbourhoods and made it more difficult for black people to obtain mortgages. Land was cheaper at the urban periphery, and suburban housing was mostly composed of single-family, owner-occupied homes. Since black people generally lacked the financial resources to move there, residential segregation continued and peaked around 1970, though some rich African Americans with high levels of education and income were able to achieve home ownership after 1940. These trends are collectively the most suitable candidate to act as a crucial factor of subjective preferences and individual adaptive expectations that this study's estimations identify as being extremely important since 1970. In 1968, the US Congress moved to create a secondary market for mortgages. In particular, race- and gender-based discrimination in the mortgage market emerged as a key issue in the early 1970s, which caused the passage of the Equal Credit Opportunity Act. The type of situation experienced in those years replicated the one studied in KOLLMANN and FISHBACK (2011), who named mortgage practices in the 1920s and 1930s responsible for the spatial distribution of blacks and whites across territory.

In this context, lower land prices far from CBDs have favoured the construction of larger, single-family detached dwellings (CASE, 1986). As BOUSTAN and MARGO (2013) discuss in relation to the rest of the United States, Massachusetts experienced strong white suburbanisation during the postwar period through increased home ownership rates for black people who remained in urban centres.

7 Conclusions

This study has proposed a Bayesian approach to track the importance of distance from Boston in shaping Massachusetts' population density distribution. By studying census

data across more than 80 years, it is possible to identify an interesting evolutionary pattern. Distance is a key element in determining location preferences, though its impact on the distribution function does not decrease in tandem with the improvement of transportation infrastructure, namely across time, given the regular investment plans put in place by the state. Instead, the coefficient associated with distance reveals an inverted U shape whose magnitude varies according to changes in the importance of an area's racial composition among citizens' preferences for residential areas. Put differently, ease of access to Boston has not always been the dominant component of subjective preferences; institutional setting may help to reinforce the ethnic dimension in the composition of citizens' residential preferences.

The idea of adopting a measure of subjective distance instead of a simple spatial distance variable underscores important empirical figures that may provide interesting suggestions for policy. Regional planning targeting the improvement of infrastructural quality to boost accessibility to CBDs and shape population density distribution may be totally ineffective, or at least lose effectiveness, if the location choices are mostly driven by other factors such as ethnic preferences. Instead, public authorities should consider the potential consequences of segregation patterns that an improved transportation system may induce. Of course, these results would be more precise if a longer data series at the urban level were available. The technique exploited in this study would certainly allow for the identification and exploration of the creation of urban ghettos and for tracking their evolution over time.

Lastly, this study has assumed that population densities are independent over time. The only relationship between them is produced by the historical prior in only slight way. Actually, a Bayesian dynamic space-time gamma model with dynamic gamma spatial frailties and the vector of regression coefficients modelled a priori as a vector random walk is in progress.

References

- BOUSTAN LP, MARGO R (2013) A Silver Lining to White Flight? White Suburbanization and African-American Homeownership, 1940-1980. *Journal of Urban Economics* 78: 71-80
- CASE K (1986) The market for Single-Family Homes in the Boston Area. *New England Economic Review*, Federal Reserve Bank of Boston, issue May: 38-48
- COLLINS WJ, MARGO RA R (2011) Race and Home Ownership from the End of the Civil War to the present. *American Economic Review: P&P* 101(3): 355-359
- CUTLER DM, GLAESER E, VIGDOR L (1999) The Rise and Decline of the American. *Journal of Political Economy* 107: 455-506
- DURLAUF S (2004) Neighborhood Effects. In: HENDERSON JV, THISSE JF (Eds) *Handbook of Regional and Urban Economics vol.4* Ch.50: 2173-2242
- EPIFANI I, NICOLINI R (2013) On the density distribution across space: a probabilistic approach. *Journal of Regional Science* 53(3): 481-510
- FETTER DK (2012) The 20th-century increase in the US home ownership: facts and hypotheses. In: Fishback P, Snowden K, White E (Eds) *Housing and Mortgage Markets in Historical Perspective* Chicago University Press
- FIRTH D (1988) Multiplicative errors: Log-normal or gamma?. *Journal of the Royal Statistical Society: Series B* 50: 266-268
- GERARDI KS, ROSEN HS, WILLEN PS (2010) The impact of Deregulation and Financial Innovation on Consumers: The Case of the Mortgage Market. *Journal of Finance* 65(1): 333-360
- GEWEKE J (1993) Bayesian Treatment of the Independent Student t Linear Model. *Journal of Applied Econometrics* 8: 19-40
- GLAESER E (2005) Reinventing Boston: 1630-2003. *Journal of Economic Geography* 5: 119-153

- GLAESER E, WARD BA (2009) The causes and consequences of land use regulation: evidence from the Greater Boston. *Journal of Urban Economics* 65(3): 265-278
- HEINTZ J, POLLIN R, GARRET-PELTIER H (2009) *How infrastructure Investments Support the U.S. Economy: Employment, Productivity and Growth*. Political Economy Research Institute
- HELSLEY RW, STRANGE, WC (2007) Urban interactions and spatial structure. *Journal of Economic Geography* 7: 119-138
- IHS Global Insight (2012) *US Metro Economies*. Acta for The United States Conferences of Mayors and the Council on Metro Economies in the New American City, July 2012
- KOLLMAN TM, FISHBACK PV (2011) The New Deal, Race and Home Ownership in the 1920s and 1930s. *American Economic Review: P&P* 101(3): 366-370
- LeSAGE JP, PACE KR (2009) *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, FL
- McCULLAGH P, NELDER JA (1989) *Generalized Linear Models*. Chapman and Hall, London
- NAIRN AGM, O'NEILL GJ (1988) Population Density Functions: A differential Equation Approach. *Journal of Regional Science* 28: 89-102
- PARR J (1985a) A Population-Density Approach to Regional Spatial Structure. *Urban Studies* 22: 289-303
- PARR J (1985b) The Form of the Regional Density Function. *Regional Studies* 19: 535-546
- PLUMMER M (2013) *JAGS Release 3.4.0: Just Another Gibbs Sampler* URL <http://mcmc-jags.sourceforge.net/>
- QUIGLEY JM (1985) Consumer Choice of Dwelling, Neighborhood and Public Services. *Regional Science and Urban Economics* 15: 41-63
- R CORE TEAM (2014) *R: A Language and Environment for Statistical Computing*. URL <http://www.R-project.org/>, R foundation for Statistical Computing, Vienna, Austria

ROSS SL, TURNER MA (2005) Housing Discrimination in Metropolitan America: Explaining Changes between 1989 and 2000. *Social Problems* 52(2): 152-180

SIMCOX AC (1992) *Water Resources of Massachusetts. Water Resources Investigation Report 90-4144*, US Geological Survey, Boston, MA

SONG SH (1996) Some Tests of Alternative Accessibility: A Population Density Approach. *Land Economics* 72: 474-482

TOPA G, ZENOU Y (2015) Neighbourhood Effects versus Network Effects. In: DURANTON G, HENDERSON JV, STRANGE W (Eds) *Handbook of Regional and Urban Economics vol.5*, 561-624, Elsevier Publisher, Amsterdam.

Prior	<i>BM</i>				<i>MEC</i>				<i>MFI</i>	
	<i>DIC</i>		Outliers		<i>DIC</i>		Outliers		<i>DIC</i>	Outliers
	diffuse	historical	diffuse	historical	diffuse	historical	diffuse	historical	Diffuse	Diffuse
1880	-218.0	NA	5.5%	NA	NA	NA	NA	NA	NA	NA
1890	-141.4	-142.4	4.1%	4.1%	NA	NA	NA	NA	NA	NA
1930	59.7	NA	4.8%	NA	63.4	NA	4.6%	NA	66.1	4.6%
1940	84.7	80.1	4.3%	4.8%	88.5	83.3	4.3%	4.0%	90.6	4.3%
1950	157.8	151.9	4.3%	4.3%	NA	NA	NA	NA	NA	NA
1960	271.8	264.89	3.1%	3.4%	277.4	269.4	3.4%	3.1%	279.5	3.4%
1970	367.1	359.3	2.8%	2.9%	373.5	361.5	2.9%	2.9%	375.4	2.9%
1980	397.7	387.8	3.1%	3.4%	363.6	386.7	2.9%	2.6%	365.4	2.9%
1990	439.7	428.6	3.1%	3.1%	369.2	385.8	2.6%	3.1%	366.4	2.9%
2000	462.1	451.2	3.4%	3.4%	352.2	368.4	2.6%	2.9%	343.4	2.6%
2010	481.2	471.0	3.4%	3.1%	351.7	365.6	2.9%	2.9%	326.7	3.4%

Table 1: DIC values and percentage of Bayesian outliers in *BM*, *MEC* and *MFI*. Cf. Appendix C for more details on the Bayesian outliers. Legend: NA: Not available data.

	1880	1890	1930	1940	1950	1960	1970	1980	1990	2000	2010
β_0 (Intercept)	-1.53 (0.21) (-1.98, -1.13)	-1.461 (0.24) (-1.97, -1.04)	-1.24 (0.31) (-1.91, -0.71)	-1.22 (0.39) (-1.88, -0.71)	-1.14 (0.29) (-1.79, -0.63)	-0.94 (0.27) (-1.51, -0.46)	-0.78 (0.26) (-1.32, -0.31)	-0.65 (0.22) (-1.12, -0.24)	-0.53 (0.21) (-0.98, -0.14)	-0.48 (0.21) (-0.91, -0.10)	-0.43 (0.21) (-0.88, -0.04)
β_3 (Dist)	-1.01 (0.18) (-1.38, -0.67)	-1.24 (0.20) (-1.66, -0.86)	-1.70 (0.25) (-2.22, -1.23)	-1.73 (0.25) (-2.24, -1.27)	-1.80 (0.24) (-2.29, -1.33)	-1.77 (0.23) (-2.24, -1.33)	-1.75 (0.21) (-2.19, -1.34)	-1.64 (0.20) (-2.04, -1.27)	-1.53 (0.18) (-1.90, -1.19)	-1.55 (0.17) (-1.91, -1.23)	-1.54 (0.18) (-1.90, -1.21)
$\alpha^{(t)}$	2.59 (1.21) (0.93, 5.60)	2.03 (0.91) (0.78, 4.32)	1.33 (0.57) (0.51, 2.76)	1.36 (0.59) (0.53, 2.79)	1.32 (0.56) (0.52, 2.68)	1.53 (0.67) (0.60, 3.20)	1.67 (0.72) (0.65, 3.38)	1.95 (0.85) (0.75, 4.02)	2.36 (1.06) (0.90, 4.92)	2.29 (1.02) (0.88, 4.76)	2.40 (1.11) (0.87, 5.20)
θ	0.84 (0.07) (0.72, 0.98)	0.82 (0.06) (0.71, 0.95)	0.64 (0.04) (0.56, 0.72)	0.66 (0.04) (0.58, 0.74)	0.70 (0.05) (0.62, 0.80)	0.80 (0.05) (0.70, 0.90)	0.90 (0.06) (0.79, 1.02)	1.03 (0.07) (0.90, 1.18)	1.10 (0.08) (0.96, 1.25)	1.19 (0.08) (1.04, 1.35)	1.20 (0.08) (1.04, 1.38)

(a) *BM*

	1930	1940	1950	1960	1970	1980	1990	2000	2010
β_0 (Intercept)	-1.26 (0.36) (-2.07, -0.65)	-1.27 (0.39) (-2.11, -0.62)	-0.98 (0.34) (-1.72, -0.41)	-0.78 (0.30) (-1.46, -0.23)	-0.66 (0.25) (-1.21, -0.23)	-0.58 (0.22) (-1.08, -0.18)	-0.54 (0.19) (-0.94, -0.19)	-0.50 (0.18) (-0.88, -0.16)	
β_1 (Z)	-0.11 (0.27) (-0.71, 0.35)	-0.10 (0.27) (-0.70, 0.36)	-0.11 (0.24) (-0.63, 0.30)	-0.13 (0.22) (-0.60, 0.25)	-0.17 (0.17) (-0.55, 0.14)	-0.11 (0.17) (-0.48, 0.17)	-0.09 (0.13) (-0.37, 0.14)	-0.05 (0.13) (-0.34, 0.17)	
β_2 ($Z \times Dist$)	0.18 (0.29) (-0.31, 0.82)	0.20 (0.29) (-0.31, 0.82)	0.13 (0.25) (-0.29, 0.68)	0.12 (0.23) (-0.28, 0.63)	0.04 (0.19) (-0.29, 0.46)	0.16 (0.19) (-0.16, 0.61)	0.03 (0.15) (-0.22, 0.34)	0.07 (0.15) (-0.18, 0.41)	
β_3 (Dist)	-1.73 (0.27) (-2.26, -1.23)	-1.79 (0.27) (-2.34, -1.27)	-1.84 (0.25) (-2.33, -1.35)	-1.79 (0.23) (-2.26, -1.33)	-1.56 (0.20) (-1.96, -1.18)	-1.39 (0.18) (-1.75, -1.05)	-1.26 (0.16) (-1.58, -0.97)	-1.15 (0.15) (-1.46, -0.89)	
β_4 (<i>Mix</i>)	0.12 (0.10) (-0.10, 0.29)	0.12 (0.10) (-0.11, 0.27)	-0.02 (0.13) (-0.32, 0.19)	0.00 (0.10) (-0.23, 0.17)	-0.42 (0.08) (-0.57, -0.28)	-0.49 (0.07) (-0.63, -0.37)	-0.54 (0.06) (-0.66, -0.43)	-0.60 (0.06) (-0.72, -0.49)	
$\alpha^{(t)}$	1.16 (0.59) (0.41, 2.73)	1.12 (0.58) (0.41, 2.64)	1.27 (0.63) (0.46, 2.92)	1.43 (0.74) (0.51, 3.35)	2.12 (1.15) (0.69, 4.98)	2.60 (1.52) (0.77, 6.50)	3.54 (2.00) (1.10, 8.51)	4.02 (2.21) (1.17, 9.73)	
θ	0.64 (0.04) (0.56, 0.72)	0.66 (0.04) (0.58, 0.75)	0.80 (0.05) (0.70, 0.91)	0.90 (0.06) (0.79, 1.03)	1.14 (0.08) (0.98, 1.30)	1.31 (0.09) (1.14, 1.50)	1.55 (0.11) (1.34, 1.77)	1.64 (0.12) (1.43, 1.88)	

(b) *MEC*

Table 2: Summaries of the posterior distributions of the parameters in *Baseline model (BM)* and years $t = 1880, 1890, 1930 - 2010$ and of the parameters in *Model with Ethnic Composition (MEC)* from 1930 to 2010 (Data in 1950 for *Mix* are not available.) Statistical significant Bayesian estimations are in bold. Posterior means of the parameters are followed by (standard deviations) in the first row, 95 percent credible intervals are shown in the second row and Bayesian (posterior predictive) p -value only for the regression coefficients are in the third row. Table 2(a) refers to *BM*, and Table 2(b) to *MEC*. (Diffuse prior)

(a) *BM*

	1880	1890	1930	1940	1950	1960	1970	1980	1990	2000	2010
β_0 (Intercept)	-1.53 (0.21) (-1.98, -1.13)	-1.47 (0.15) (-1.78, -1.18)	-1.24 (0.31) (-1.91, -0.71)	-1.21 (0.21) (-1.62, -0.82)	-1.17 (0.17) (-1.50, -0.85)	-1.1 (0.15) (-1.39, -0.82)	-1.02 (0.13) (-1.29, -0.78)	-0.96 (0.12) (-1.20, -0.73)	-0.89 (0.11) (-1.12, -0.67)	-0.84 (0.11) (-1.05, -0.63)	-0.78 (0.10) (-0.98, -0.59)
β_3 (Dist)	-1.01 (0.18) (-1.38, -0.67)	-1.11 (0.13) (-1.37, -0.87)	-1.70 (0.25) (-2.22, -1.23)	-1.72 (0.17) (-2.06, -1.38)	-1.76 (0.14) (-2.03, -1.49)	-1.79 (0.12) (-2.02, -1.55)	-1.81 (0.11) (-2.01, -1.60)	-1.80 (0.09) (-1.97, -1.61)	-1.77 (0.08) (-1.93, -1.61)	-1.76 (0.08) (-1.91, -1.61)	-1.74 (0.07) (-1.88, -1.60)
α^t	2.59 (1.21) (0.93, 5.60)	2.17 (0.88) (0.93, 4.29)	1.33 (0.57) (0.51, 2.76)	1.26 (0.48) (0.55, 2.37)	1.16 (0.40) (0.54, 2.11)	1.26 (0.45) (0.57, 2.28)	1.27 (0.45) (0.58, 2.30)	1.29 (0.45) (0.59, 2.33)	1.32 (0.46) (0.61, 2.36)	1.32 (0.45) (0.62, 2.34)	1.34 (0.48) (0.61, 2.44)
θ	0.84 (0.07) (0.72, 0.98)	0.82 (0.06) (0.71, 0.94)	0.64 (0.04) (0.56, 0.72)	0.66 (0.04) (0.58, 0.74)	0.70 (0.05) (0.62, 0.80)	0.80 (0.05) (0.70, 0.91)	0.91 (0.06) (0.79, 1.03)	1.04 (0.07) (0.91, 1.18)	1.11 (0.07) (0.97, 1.26)	1.19 (0.08) (1.04, 1.35)	1.21 (0.08) (1.05, 1.38)

(b) *MEC*

	1930	1940	1950	1960	1970	1980	1990	2000	2010
β_0 (Intercept)	-1.26 (0.36) (-2.07, -0.65) 0.000	-0.13 (0.27) (-0.63, 0.41) 0.303	-0.49 (0.19) (-0.86, -0.12) 0.007	-0.55 (0.16) (-0.86, -0.26) 0.000	-0.57 (0.13) (-0.83, -0.31) 0.000	-0.57 (0.12) (-0.80, -0.35) 0.000	-0.56 (0.10) (-0.78, -0.36) 0.000	-0.56 (0.10) (-0.75, -0.37) 0.000	
β_1 (Z)	-0.11 (0.27) (-0.71, 0.35) 0.354	-0.27 (0.18) (-0.62, 0.07) 0.057	-0.24 (0.13) (-0.51, 0.00) 0.026	-0.22 (0.10) (-0.42, -0.02) 0.016	-0.20 (0.09) (-0.37, -0.02) 0.013	-0.17 (0.08) (-0.33, -0.01) 0.015	-0.15 (0.07) (-0.29, -0.02) 0.012	-0.13 (0.06) (-0.26, -0.01) 0.016	
β_2 ($Z \times Dist$)	0.18 (0.29) (-0.31, 0.82) 0.302	0.23 (0.19) (-0.13, 0.62) 0.114	0.20 (0.14) (-0.08, 0.48) 0.085	0.18 (0.12) (-0.06, 0.40) 0.061	0.16 (0.10) (-0.03, 0.36) 0.048	0.19 (0.08) (0.03, 0.36) 0.009	0.20 (0.08) (0.05, 0.34) 0.004	0.21 (0.07) (0.08, 0.35) 0.001	
β_3 (Dist)	-1.73 (0.27) (-2.26, -1.23) 0.000	-1.72 (0.17) (-2.06, -1.39) 0.000	-1.74 (0.13) (-2.00, -1.48) 0.000	-1.75 (0.11) (-1.97, -1.53) 0.000	-1.73 (0.10) (-1.92, -1.54) 0.000	-1.69 (0.08) (-1.85, -1.52) 0.000	-1.64 (0.08) (-1.79, -1.50) 0.000	-1.60 (0.07) (-1.73, -1.47) 0.000	
β_4 (<i>Mix</i>)	0.12 (0.10) (-0.10, 0.29) 0.113	0.13 (0.07) (-0.01, 0.26) 0.037	0.10 (0.06) (-0.02, 0.21) 0.046	0.08 (0.04) (-0.02, 0.16) 0.050	-0.09 (0.03) (-0.16, -0.03) 0.003	-0.20 (0.03) (-0.25, -0.15) 0.000	-0.27 (0.02) (-0.32, -0.22) 0.000	-0.32 (0.02) (-0.36, -0.28) 0.000	
α^t	1.16 (0.59) (0.41, 2.73)	0.61 (0.28) (0.22, 1.29)	1.15 (0.46) (0.48, 2.23)	1.30 (0.49) (0.58, 2.44)	1.40 (0.51) (0.63, 2.61)	1.45 (0.52) (0.65, 2.64)	1.48 (0.51) (0.68, 2.66)	1.46 (0.50) (0.69, 2.64)	
θ	0.64 (0.04) (0.56, 0.72)	0.66 (0.04) (0.58, 0.75)	0.80 (0.05) (0.70, 0.90)	0.90 (0.06) (0.78, 1.03)	1.08 (0.07) (0.94, 1.23)	1.25 (0.09) (1.09, 1.43)	1.48 (0.10) (1.28, 1.70)	1.56 (0.11) (1.35, 1.79)	

Table 3: Summaries of the posterior distributions of the parameters in *Baseline model (BM)* and years $t = 1880, 1890, 1930 - 2010$ and of the parameters in *Model with Ethnic Composition (MEC)* from 1930 to 2010 (Data in 1950 for *Mix* are not available.) Statistical significant Bayesian estimations are in bold. Posterior means of the parameters are followed by (standard deviations) in the first row, 95 percent credible intervals are shown in the second row and Bayesian (posterior predictive) p -value only for the regression coefficients are in the third row. Table 3(a) refers to *BM*, and Table 3(b) to *MEC*. (Historical priors)

(a) *MFI*

	1930	1940	1950	1960	1970	1980	1990	2000	2010
β_0 (Intercept)	-1.27 (0.37) (-2.14,-0.64) 0.000	-1.25 (0.38) (-2.10,-0.60) 0.000	-0.97 (0.33) (-1.70,-0.40) 0.000	-0.78 (0.32) (-1.47,-0.24) 0.003	-0.66 (0.26) (-1.22,-0.21) 0.003	-0.56(0.24) (-1.07,-0.13) 0.006	-0.50(0.19) (-0.90,-0.14) 0.005	-0.41 (0.19) (-0.82,-0.05) 0.012	
β_1 (<i>Z</i>)	-0.12 (0.28) (-0.74,0.35) 0.356	-0.10 (0.27) (-0.68,0.37) 0.373	-0.11 (0.23) (-0.63,0.32) 0.323	-0.12 (0.22) (-0.61,0.26) 0.292	-0.17 (0.18) (-0.57,0.14) 0.141	-0.12 (0.18) (-0.53,0.18) 0.248	-0.11 (0.13) (-0.40,0.12) 0.197	-0.08 (0.13) (-0.38,0.14) 0.268	
β_2 (<i>Z</i> × <i>Dist</i>)	0.18 (0.29) (-0.32,0.81) 0.285	0.20 (0.29) (-0.30,0.83) 0.263	0.13 (0.24) (-0.29,0.66) 0.319	0.12 (0.24) (-0.28,0.65) 0.328	0.03 (0.20) (-0.31,0.48) 0.480	0.17 (0.20) (-0.15,0.62) 0.198	0.00 (0.15) (-0.25,0.33) 0.466	0.04 (0.15) (-0.21,0.39) 0.434	
β_3 (<i>Dist</i>)	-1.74 (0.27) (-2.29,-1.23) 0.000	-1.81 (0.27) (-2.36,-1.29) 0.000	-1.83 (0.25) (-2.35,-1.35) 0.000	-1.79 (0.23) (-2.26,-1.33) 0.000	-1.56 (0.20) (-1.96,-1.18) 0.000	-1.40 (0.18) (-1.76,-1.06) 0.000	-1.26 (0.15) (-1.57,-0.97) 0.000	-1.15 (0.15) (-1.45,-0.87) 0.000	
β_4 (<i>Mix</i>)	0.10 (0.11) (-0.14,0.30) 0.172	0.06 (0.14) (-0.24,0.31) 0.318	-0.04 (0.14) (-0.33,0.21) 0.405	-0.01 (0.10) (-0.24,0.16) 0.478	-0.44 (0.08) (-0.60,-0.30) 0.000	-0.57 (0.08) (-0.74,-0.42) 0.000	-0.68 (0.08) (-0.83,-0.53) 0.000	-0.84 (0.08) (-1.00,-0.69) 0.000	
β_5 (<i>Mix</i> × <i>Dist</i>)	0.01 (0.20) (-0.38,0.41) 0.497	0.11 (0.27) (-0.44,0.66) 0.336	0.01 (0.19) (-0.37,0.38) 0.484	-0.08 (0.16) (-0.39,0.25) 0.314	-0.08 (0.08) (-0.24,0.08) 0.150	-0.17 (0.08) (-0.33,-0.02) 0.012	-0.23 (0.07) (-0.37,-0.09) 0.000	-0.36 (0.07) (-0.51,-0.22) 0.000	
$\alpha^{(9)}$	1.15 (0.60) (0.41,2.68)	1.11 (0.56) (0.40,2.53)	1.27 (0.65) (0.46,2.88)	1.41 (0.72) (0.50,3.17)	2.13 (1.22) (0.68,5.33)	2.40 (1.38) (0.73,5.89)	3.29 (1.77) (1.01,7.72)	3.66 (2.06) (1.09, 8.97)	
θ	0.64 (0.04) (0.55,0.72)	0.66 (0.04) (0.58,0.75)	0.80 (0.05) (0.70,0.91)	0.90 (0.06) (0.79,1.03)	1.13 (0.08) (0.98,1.29)	1.33 (0.09) (1.16,1.52)	1.59 (0.11) (1.38,1.82)	1.75 (0.12) (1.52,2.00)	

Table 4: Summaries of the posterior distributions of the parameters in *Model with full interactions (MFI)* from 1930 to 2010 (Data in 1950 for *Mix* are not available.) Statistical significant Bayesian estimations are in bold. Posterior means are followed by (standard deviations) in the first row, 95 percent credible intervals are in the second row and Bayesian (posterior predictive) *p*-value only for the regression coefficients are shown in the third row. (Diffuse prior)

(a) *BM*

	1880	1890	1930	1940	1950	1960	1970	1980	1990	2000	2010
Suffolk	0.29	0.28	0.30	0.33	0.37	0.45	0.52	0.56	0.54	0.52	0.53
Franklin	1.30	1.12	0.76	0.76	0.71	0.84	0.91	1.05	1.20	1.23	1.31
Playmouth	2.39	2.47	3.14	3.13	3.12	2.79	2.46	2.20	2.12	2.10	2.11
Middlesex	0.88	0.86	0.90	0.92	1.00	1.05	1.12	1.16	1.14	1.16	1.18
Bristol	0.80	0.62	0.57	0.59	0.61	0.71	0.75	0.80	0.84	0.88	0.90
Berkshire	0.37	0.24	0.10	0.10	0.08	0.10	0.11	0.15	0.22	0.23	0.26
Hampden	0.41	0.30	0.12	0.12	0.11	0.11	0.12	0.16	0.19	0.20	0.20
Essex	0.68	0.69	0.76	0.80	0.86	0.96	1.02	1.06	1.04	1.05	1.06
Hampshire	0.69	0.62	0.41	0.42	0.35	0.37	0.38	0.42	0.50	0.50	0.51
Dukes	0.79	0.91	1.04	0.93	1.25	1.10	1.24	1.39	1.03	1.17	0.85
Worcester	1.59	1.38	1.29	1.30	1.33	1.48	1.57	1.71	1.69	1.70	1.69
Norfolk	1.97	2.45	2.46	2.42	2.40	2.14	2.07	2.04	1.97	2.01	2.01
Barnstable	0.99	1.25	1.51	1.37	1.22	1.11	0.92	0.71	0.70	0.63	0.69
Nantucket	0.88	0.86	0.69	0.70	0.65	0.74	0.78	0.76	0.82	0.65	0.66

(b) *MEC*

	1930	1940	1950	1960	1970	1980	1990	2000	2010
Suffolk	0.25	0.27		0.35	0.39	0.50	0.50	0.67	0.68
Franklin	0.75	0.70		0.79	0.91	1.08	1.13	1.41	1.42
Playmouth	2.70	2.77		2.58	2.28	2.04	1.70	1.50	1.33
Middlesex	1.18	1.23		1.32	1.40	1.34	1.44	1.34	1.38
Bristol	0.55	0.55		0.68	0.74	0.87	0.83	0.86	0.83
Berkshire	0.10	0.08		0.09	0.11	0.19	0.23	0.35	0.41
Hampden	0.12	0.11		0.11	0.12	0.24	0.28	0.36	0.37
Essex	0.68	0.71		0.85	0.91	0.93	0.91	0.88	0.84
Hampshire	0.41	0.38		0.35	0.38	0.48	0.57	0.61	0.68
Dukes	1.10	1.07		1.11	1.17	1.26	1.20	1.23	1.19
Worcester	1.43	1.41		1.57	1.74	1.77	1.76	1.67	1.58
Norfolk	2.62	2.62		2.25	2.16	1.88	1.81	1.66	1.70
Barnstable	1.33	1.18		1.00	0.86	0.64	0.65	0.57	0.65
Nantucket	0.77	0.79		0.78	0.82	0.73	1.00	0.89	0.96

(c) *MFI*

	1930	1940	1950	1960	1970	1980	1990	2000	2010
Suffolk	0.12	0.12		0.22	0.24	0.39	0.35	0.56	0.51
Franklin	0.65	0.62		0.71	0.84	1.02	1.04	1.32	1.29
Playmouth	2.55	2.71		2.48	2.17	1.95	1.71	1.55	1.43
Middlesex	1.02	1.07		1.17	1.24	1.23	1.34	1.21	1.25
Bristol	0.51	0.53		0.65	0.71	0.84	0.83	0.88	0.86
Berkshire	0.07	0.06		0.07	0.08	0.16	0.18	0.29	0.32
Hampden	0.10	0.10		0.10	0.11	0.24	0.27	0.37	0.38
Essex	0.64	0.66		0.82	0.89	0.90	0.90	0.89	0.89
Hampshire	0.35	0.33		0.31	0.34	0.45	0.54	0.56	0.65
Dukes	0.80	0.79		0.87	0.96	1.10	1.04	1.10	1.10
Worcester	1.33	1.33		1.50	1.65	1.72	1.76	1.65	1.63
Norfolk	2.40	2.42		2.10	2.05	1.78	1.73	1.59	1.63
Barnstable	1.23	1.11		0.96	0.80	0.60	0.60	0.53	0.58
Nantucket	0.55	0.56		0.60	0.66	0.62	0.88	0.85	0.93

Table 5: Temporal evolution of the posterior means of the county gamma frailties under diffuse priors and *BM* (with only distance) for decades 1880, 1890 and from 1930 to 2010 in Table 5(a); under diffuse priors and *MEC* (with *Dist*, *Mix* and *Z* land) from 1930 to 2010 in Table 5(b); under diffuse priors and *MFI* (with *Dist*, *Mix*, *Z* land and the interaction *Dist* and *Mix*) from 1930 to 2010 in Table 5(c). Legend: the color blue denotes counties with mean frailty “much” less than one and color yellow is used to highlight the counties with mean frailty “much” greater than one. Lastly, Barnstable is green to denote a county effect changing over time.

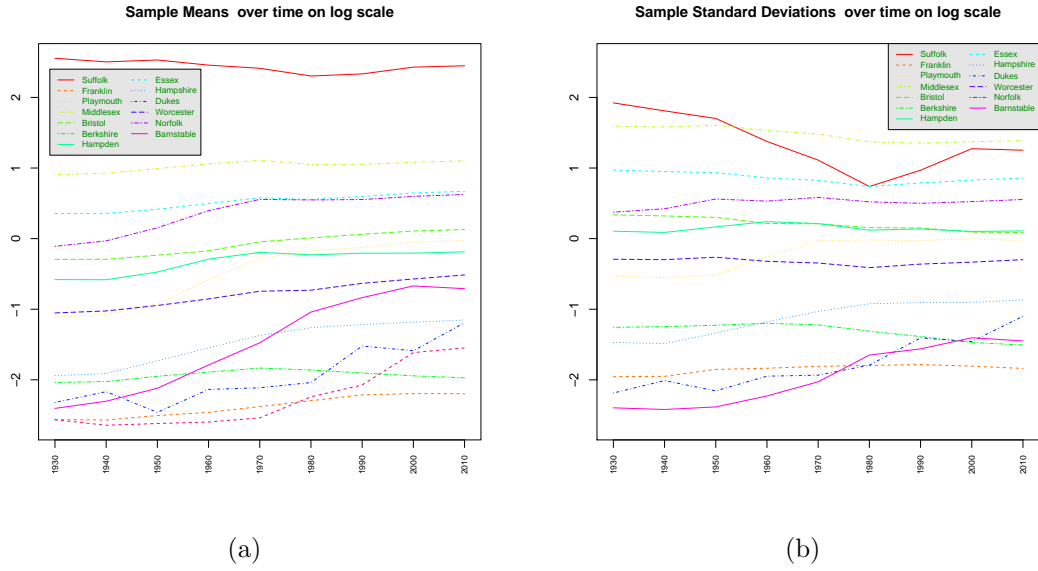


Figure 1: Sample means -in Figure 1(a)- and sample standard deviations -in Figure 1(b)- of the municipality population densities per square mile of every county in Massachusetts, on a logarithmic scale, from 1930 to 2010. The values of sample means and standard deviations on original scale are shown in Appendix A, Table 6. Source: US Bureau Census, Calculus: authors.

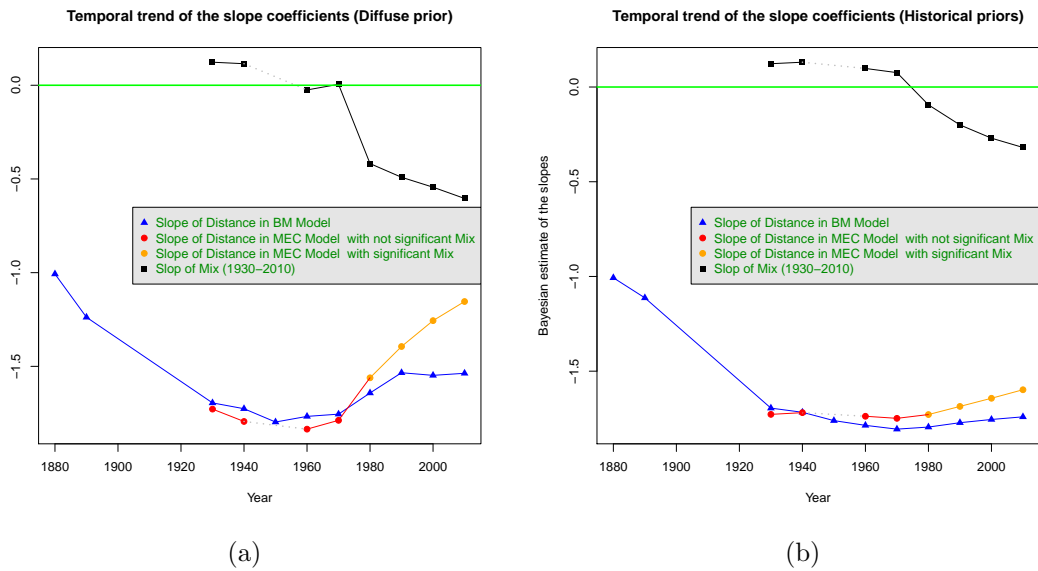


Figure 2: Bayesian estimates (posterior means) of the slope coefficients of *Dist* in *Baseline model* (*BM*) for the years 1880, 1890, 1930-2010 and of *Dist* and *Mix* for *Model with Ethnic Composition* (*MEC*) from 1930 to 2010. Data in 1950 for *Mix* are not available. Legend: blue up-pointing triangles: coefficient of *Dist* in *BM*; solid black squared: coefficient of *Mix* in *MEC*; solid red and orange circles: coefficient of *Dist* in *MEC*. In Figure 2(a) we report the outcomes in the case of diffuse prior and in Figure 2(b) the outcomes in the case of historical priors.