This is the post-print version of the article: Epifani, I., Ghiringhelli, C., Nicolini, R. (2020): "Population distribution over time: modelling local spatial dependence with a CAR process" Spatial Economic Analysis, vol. 15(2), pp. 120-144. https://doi.org/10.1080/17421772.2020.170844

Title Population distribution over time: modeling local spatial dependence with a CAR process

Ilenia Epifani Politecnico di Milano, Dip. di Matematica, P.zza L. da Vinci, 32, I-20133 Milano (Italy), E-mail: ilenia.epifani@polimi.it

Chiara Ghiringhelli Università della Svizzera Italiana, Dip. di Econometria, Via G. Buffi, 13, Lugano, (Switzerland), E-mail: chiara.ghiringhelli@usi.ch.

Rosella Nicolini (Corresponding author) Departament d'Economia Aplicada, Universitat Autònoma de Barcelona, Edifici B – Campus UAB, 08193 Bellaterra (Spain), E-mail: rosella.Nicolini@uab.cat.

Acknowledgements We are grateful to the editor P. Elhorst and two anonymous reviewers for insightful suggestions as well as the participants to NARSC 2016 (Portland) and XII SEA 2018 (Wien) for interesting comments. We also acknowledge the information provided by officers at CIS Mass-State, NAco and Massachusetts Archives. All errors are our own responsibility. This work was supported by the Ministerio de Economia, Industria y Competitividad under Grant ECO2014-52506-R, Generalitat de Catalunya 2017SGR207 and XREPP.

Population distribution over time: modeling local spatial dependence with a CAR process

Running Title Population distribution and spatial dependence

Abstract

We investigate the effectiveness of local spatial dependence in shaping the population density distribution. We model individual location preferences by considering status-related features of a given spatial unit and its neighbors as well as local random spatial dependence. Our novelty is framing such a dependence through CAR (*Conditionally Autoregressive*) census random effects that we add to an SLX (*Spatially Lagged Explanatory Variable- X-*) setting. Our results not only confirm that controlling for the spatial dimension is relevant but also indicate that local spatial dependence warrants consideration in determining the population distribution of recent decades. In this respect, our framework turns to be useful for the analysis of microdata in which individual relationships (in a same spatial unit) enforce local spatial dependence.

Keywords: Hierarchical Bayesian spatio-temporal model, Population density distribution, Spatial conditionally autoregressive (CAR) model, Spatial interaction.

JEL Codes: C11, N92, R19; *AMS Classification:* 62F15, 62P20, 91G70.

1 Introduction

Individual location choices are often driven by features that define the attractiveness of a neighborhood. Economic scholars have conceived the notion of *neighborhood status*, referring to a group of socioeconomic features for a given spatial unit (Rosenthal and Ross, 2015). By extension, the spatial association of environmental characteristics factors into individuals' preferences, but decisions about where to settle are subject to individuals' budget constraints (see Topa and Zenou, 2015). However, the relevance of those attributes is not limited to a given period: the consolidation of specific features of a neighborhood in different periods affects individuals' preferences by factoring into the neighborhood's reputation (for instance, a neighborhood's reputation as a ghetto). In addition, the spatial association of environmental features is not always exclusive to one spatial unit but often spans across adjacent units, and hence this creates local spatial dependence among them.

In this study, we aim at developing an empirical research by assessing the importance of local spatial dependence in shaping the population density distribution. Our analysis is performed with the aim of providing evidence for the case of Massachusetts from 1970 to 2010.

We elaborate an original empirical setting of analysis grounded in spatial econometrics and follow a Bayesian approach to achieve our goal. Two motivations drive our choice. First, an econometric model, designed to capture complex spatio-temporal relationships in a sample of heterogeneous data, requires the consideration of several parameters and poses the risk of overparameterization. The potential problem of overfitting calls for applying a Bayesian approach to make inferences, which, by using prior distributions, allows for prior constraints on the parameters. Second, we recognize the opportunity to exploit the logic of the Bayesian treatment of model uncertainty to narrow the choice between global and local dependence in the empirical modelings as argued in LeSage (2014). Along these lines, our strategy involves first defining a baseline model without spatial dependence and, later on, augmenting it by including the most suitable spatial effects. Spatial dependence in our study includes both spatially lagged covariates and spatial dependence in the disturbances.

Our baseline model builds on a lognormal function that is suitable for modeling a monocentric distribution pattern that includes the most relevant variables for individuals' location choices. Next, we introduce the local spatial interaction (at the census tract level). For one, canonical variables for shaping the population density distribution (such as income or education in Glaeser, 2008) often involve temporal and spatial relationships, and spatial correlation patterns are well described by augmenting the baseline model with spatially lagged components.

As a novelty for this type of setting, to model spatial dependence commonly detected in small units, we formalize our analysis by applying a Bayesian CAR (*Conditionally Autoregressive*) process at the census tract level in the disturbance terms. A CAR model is often suitable when data are characterized by a spatio-temporal structure. In the CAR model, the probabilistic explanation for a variable at a given location depends on the average of neighboring values, meaning that the model emphasizes the importance of the local contagion effect. In the literature, De Oliveira (2012) proposes a method for developing a Bayesian analysis of hierarchical models in a CAR setting. Previously, Arab et al. (2008) demonstrated that a CAR process, implemented in a Bayesian hierarchical setting, is more suitable than traditional covariance-based methods because the former allows complicated structures to be modeled with different hierarchical levels.

Our empirical strategy also entails important consequences for modeling time dependence in the setting that we investigate, since the time dimension is essentially subsumed by any feasible dynamic specification strategy for priors. On the one hand, we model a priori autoregressive regression coefficients so that they embed a memory process into their dynamics. On the other hand, spatial-time dependence is integrated into county random frailties when we consider population density as a function of all county random effects, in both current and earlier times.

Regarding our data choice, new releases of US Census data allow for the development of a robust empirical framework at the finest spatial level, and newly released version(s) of the American Communitarian Survey complement the missing information. With those data, we construct a database covering the period from 1970 to 2010, which allows us to trace our source of information for at least 40 years and, according to the data released in the NHGIS project, at the census tract level.[1]

Our results confirm that land organization in Massachusetts has accommodated a monocentric-type structure, since the estimated coefficient of the distance from Boston, qualified as a *central business district* (CBD), is always negative and statistically different from zero, meaning that the proximity to Boston is a truly dominant factor for individual location choices.

As for the economic predictions of our model, local covariates such as income, level of education, and ethnic composition that are proxies for the status of a neighborhood are always either significant or not, regardless of the presence of any kind of random frailties. Those predictions emphasize that segregation effects have worsened over time, mostly because of ethnic or income discrimination.

At the spatial level, our estimations also reveal an interesting interplay between county and census random effects (once spatial interaction has been taken into account): whereas the former control for heteroscedasticity (i.e., different variability) across the counties (above all, the counties farthest from

[1] Unfortunately, we cannot treat available census tract data as panel data because the number and structure of census tracts are not constant over time.

Boston have the largest variances), the latter are the chief drivers of heterogeneity.

On the whole, our estimates point to the growing importance of local spatial dependence over time: the way in which we model spatial dependence makes proximity between census tracts crucial in shaping the population density distribution. In perspective, this result highlights that today's society is concerned with reducing spatial and communication distances: all of these efforts cause distance to shrink and land structure to reorganize. Hence, the local-spatial dimension acquires a relevant role in individual location choices because agents are concerned with environmental features that they can find at a local level, and this drives their decisions. This type of bottom-up approach allows the small spatial dimension to overcome the relevance of other features brought at a higher spatial level.

The remainder of this paper is organized as follows. Section 2 proposes a literature review. Section 3 sketches the theoretical framework of our analysis, after which we outline our data in Section 4 and introduce the hierarchical Bayesian spatio-temporal lognormal model in Section 5. Next, we describe the goodness-of-fit and model selection in Section 6, present the results of estimations in Section 7, and offer our conclusions in Section 8. Additional statistics, robustness checks and details about the implementation of estimations appear in Appendices A-D.

2 Literature review

The interest in spatial dependence in understanding the evolution of socioeconomic phenomena in social science has consolidated over time (Anselin, 2003). In this paper, we examine the interplay between the relevance of neighborhood status factors and local spatial dependence as determinants of the population density distribution. In doing so, we seek to define a framework that allows us to account for how certain location determinants can influence individuals' location decisions. Of course, the subject of our study is nothing new, for economics scholars have long devoted attention to elucidating population growth in urban environments. Among them, Henderson (1988) has argued that population growth in cities is associated with residents' level of education, and Glaeser (2008) has described additional factors, including amenities, housing, crime, and transport infrastructure, that can drive individuals' decisions about where to settle in cities. Referring to the United States, Carruthers and Mulligan (2008) have stressed that features affecting population changes in urban areas differ from those in rural areas because individuals might be driven by different incentives or preferences. For example, Boarnet et al. (2005) have shown that the level of spending on education positively affects population density, and changes in population density in a given area influence the rate of similar changes in surrounding areas as well. Although they examine different settings, Boarnet et al. (2005), Carruthers and Mulligan (2008), and Chi et al. (2011) have all assessed the existence of severe spatial dependence between small spatial units (municipalities). Among other advances in that line of research, one of the most innovative contributions is by Costa da Silva et al. (2017), which overcomes limitations in identifying determinants of rural and urban population growth in Brazil by providing an analysis that takes local spatial dependence into account. As their results suggest, determinants of population dynamics in one spatial unit tend to affect population dynamics in surrounding units as well.

To further complement that line of research, we study spatial land organization in Massachusetts (US) according to *census tract* data. By modeling determinants of individuals' location decisions in both rural and urban environments, our research strategy links to the urban population growth literature, but we add a further framework to take into account the spatial land structure of the state. As Epifani and Nicolini (2013, 2017) have argued, a clear technical advantage of analyzing census tract data, particularly for Massachusetts, is that the state's geographical structure favors the adoption of a monocentric framework of analysis because Boston has consolidated its territorial attractiveness (for all of the state) and qualifies as a CBD.

Nevertheless, the effects of spatial land organization cannot be dissociated from the neighborhood relationships stemming from the interaction among individuals. That type of relationship (or spatial correlation) between individuals belonging to two spatial units tends to decrease as the distance between these units increase. So, those dynamics can be adequately represented by using a spatial autoregressive framework in which changes in a variable of interest in a given unit depend on changes in the same variable in other spatial units. Accordingly, we model spatial proximity by introducing a contiguity matrix, discussed in Section 3. A reasonable assumption is that a territorial unit is more likely to display a feature (e.g., residents' level of education) if the neighboring units also display that feature. As such, the extent of the influence of the other spatial units on the given unit can be controlled with an adjacency matrix that takes each unit's spatial contiguity into account.

In spatial econometrics, researchers often model local dependence (or local spillovers) by referring to the SDEM (*Spatial Durbin Error model*) in which the errors exhibit a SAR (*Spatial Autoregressive*) structure. According to Elhorst (2014), SDEMs are convenient for the estimation of direct and indirect local effects; the regression coefficients of the covariates belonging to a given spatial unit indicate the direct effects, whereas the indirect effects correspond to the regression coefficients of spatially lagged covariates. However, Arab et al. (2008) have additionally argued that a CAR model is more flexible than a SAR model, for the latter introduces an explicit spatial autocorrelation structure strongly connected with the weighted spatial matrix. From a probabilistic point of view, Ver Hoef et al. (2018a) show that any SAR model can be written as a unique CAR model and that any CAR model can be written as a non-unique SAR model. From an applied perspective, Wall (2004) emphasizes that empirical models are likely to include discrete information at a neighborhood level modeled as a CAR or SAR process. In particular, the adoption of CAR models has been quite extensive in public health studies: Jin et al. (2005) conclude such that the CAR framework is appropriate for mapping the spreading of a disease over geographical units such as counties, census tracts, or zip codes.

Our strategy is inspired by the previous arguments and has its roots in the argument put forward by Ver Hoef et al. (2018b). According to their study, CAR and SAR models are designed to model local spatial dependence, since they are network-based models linked to agents' connectivity in ecological data. In this spirit, we aim to transpose that idea in a different setting (namely, dealing with individuals' location choices) but seek to preserve the concern to take into account potential individual connections or contacts by means of connectivity via local dependence. Hence, our choice is to introduce a CAR structure into our empirical framework. A challenge preventing the extensive use of CAR models is their limitation in developing a proper functional form to define adequate priors. We overcome that limitation by imposing the CAR structure not on data but on random effects in order to preserve the local spatial dependence of our variable of interest - that is, the population density distribution- among spatial units while avoiding the problems associated with selecting an appropriate functional form.

3 The setting

From a theoretical perspective, our setting of analysis builds on the spatial structure frequently used in urban theory (Glaeser, 2008). Our main assumption is that individuals prefer to settle close to the CBD for several reasons. To commence modeling the spatial structure, we refer to Epifani and Nicolini's (2013) discussion about subjective preferences when presenting the monocentric spatial structure for Massachusetts; the other determinants for location are taken from the literature according to the discussion proposed in Epifani and Nicolini (2017). In our study, we also aim to introduce the dimension of local spillovers (arising from spatial proximity). Costa da Silva et al. (2017) propose an interesting theoretical framework that extends the classic models of population growth (Glaeser et al., 1995) by including spatial interaction.[2]

[2] Costa da Silva et al. (2017) propose a novel framework embedded with important features that we cannot transpose into our setting because we do not handle panel data, whereas their research relies on an aggregation of municipalities at MCA level that is comparable over time.

For formalizing our framework of analysis while retaining its generalizability, we propose that an individual's location decision is the result of a maximization problem. As in Epifani and Nicolini (2013), our underlying assumption is that Boston is a point of interest for all citizens in Massachusetts because it represents the principal economic and cultural center. As usual, this location decision is typically subject to the individual's budget constraint. At time t, an individual h chooses to settle in the spatial unit i belonging to the county g at a distance from Boston d_{it}. This decision is taken by maximizing the utility function $U(\cdot)$

$$U_{hit}(d_{it}) = U\Big(C_{hit}(d_{it})\,;\,\mathbf{X}^{(1)}_{g(i)t}\boldsymbol{\beta}^{(1)}_t\,;\,\mathbf{X}^{(2)}_{it}\boldsymbol{\beta}^{(2)}_t\,;\,\mathbf{W}_{it}\mathbf{X}^{(2)}_t\boldsymbol{\delta}_t\Big)$$

subject to the budget constraint $I_{hit} = s(d_{it}) + p_t C_{hit}(d_{it})$ for a given level of income I_{hit}. The index function $g(i)$ returns the county census tract i belongs to. Plugging the budget constraint into the utility function, we obtain

$$\max_{d_{it}} U\Big(\frac{I_{hit} - s(d_{it})}{p_t}\,;\,\mathbf{X}^{(1)}_{g(i)t}\boldsymbol{\beta}^{(1)}_t\,;\,\mathbf{X}^{(2)}_{it}\boldsymbol{\beta}^{(2)}_t\,;\,\mathbf{W}_{it}\mathbf{X}^{(2)}_t\boldsymbol{\delta}_t\Big)$$

The function U is strictly increasing in each argument, twice continuously differentiable, and strictly quasi concave. The variable $C_{hit}(d_{it})$ represents the composite good that individual h consumes at a distance d_{it} from the CBD, associated with a price index p_t. Adopting a hypothesis common in the literature, we assume that at time t, individual h in spatial unit i enjoys a level of income I_{hit} from a job in the CBD for which s(he) has to commute at cost $s \times d_{it}$ (with $s > 0$) that is proportional to the distance to the CBD. The vector $\mathbf{X}^{(1)}_{gt}$ embeds features (see Section 4) of county g, and vector $\mathbf{X}^{(2)}_{it}$ contains a few factors (e.g., level of education, per capita income, Gini index, ethnic composition) of the status of a spatial unit (here, census tract) i. The last argument, $\mathbf{W}_{it}\mathbf{X}^{(2)}_t\boldsymbol{\delta}_t$, aims to capture spillovers and refers to the average of the exogenous covariates in the neighboring census tracts. The spatial proximity of two census tracts (i, j) relies on the definition of a weight w_{ijt} arranged in a spatial weight matrix $\mathbf{W}_t$ (in the centered formulae, $\mathbf{W}_{it}$ is the ith row of $\mathbf{W}_t$ and, $\mathbf{X}^{(2)}_t$ the design matrix with the ith row $\mathbf{X}^{(2)}_{it}$). Solving the previous maximization problem, we obtain the optimal distance d^*_{it} from the CBD at which each individual decides to locate, and using that measure, we can approximate the population distribution.

In defining our setting, we make two important assumptions regarding the most convenient measure of the distance to the CBD and the structure of the spatial weight matrix $\mathbf{W}_t$. First, we approximate the optimal distance d^*_{it} from the CBD by the distance between the Boston center (one centroid) and the centroid of census tract i. We computed the Euclidean distance according to the geographical coordinates of the two centroids.[3] Second, the definition for the spatial weight matrix $\mathbf{W}_t$ is not trivial. There is a wide literature about this topic, and a complete overview can be found in Majewska (2017). In a common method, the matrix is determined according to its capacity to accommodate data (Elhorst et al., 2013). However, LeSage and Pace (2014) have reported that estimations of direct and spillovers effects are quite robust to the alternative specifications of $\mathbf{W}_t$ when it includes at least continuous or nearby spatial units (or neighborhoods). Indeed, LeSage and Pace (2014) identify a sizable degree of correlation between the number of neighborhoods considered: the more similar the magnitudes of those numbers between two alternative definitions of $\mathbf{W}_t$, the higher the correlation between spatial lags of a standard independent normal vector.[4] Hence, in the case of contiguous neighborhoods, spatial lag vectors display similar behaviors in correspondence with different specifications of $\mathbf{W}_t$ that exhibit the same type of scaling, and the inference on direct effects and spillovers is robust with respect to changes in $\mathbf{W}_t$. Therefore, in light of these insights, we consider it reasonable to settle the spatial weight matrix as a first-order row-stochastic contiguity matrix.[5]

4 Data and variables

Our statistical analysis treats spatial census tract data representing the entire state of Massachusetts during 1970-2010 and elaborated with data from the NHGIS project.[6] Our database is a time-pooled, cross-sectional collection with three dimensions: time, county, and census tract. We handle five waves of data for five decades, including data for 11 counties in 1970 and 1980 and for all 14 counties in the remaining decades. For each county, we have information about the corresponding census tracts whose number changes over time. The tract is often split into two (or more) subtracts when the size goes over the optimal limit or when the spatial territory is affected by other structural changes. Unfortunately, there is no a clear mechanism that allows for precise tracking of the changes over time.

For the period 1970-2000, data are taken directly from the US Census, and those referring to 2010 are from both the US census and the American Community Survey (ACS).

A preliminary inspection of raw population data shows that, in 1970, population density was low, and most people settled close to the largest cities. During the last 40 years, the state's territory has

[3] All geographical information and computations are run with QGIS software.

[4] For instance, when doubling the order of neighbors the correlation decreases from 0.9 to 0.7.

[5] According to our statistics, upon replacing the first-order contiguity matrix with a second-order one, the number of neighborhoods doubles (on average), and accordingly, the degree of correlation is quite high (statistics available upon request). We privilege the use of a contiguity matrix to the distance-based ones because we want to avoid the risk of overlapping two different spatial structures in the same model.

[6] Minnesota Population Center. National Historical Geographic Information System: Version 2.0. Minneapolis, MN: University of Minnesota 2011, `http://www.nhgis.org`.

progressively expanded as the population has grown. The mean and variance of the population density are almost constant over time, but the number of census tracts has increased progressively, especially those near the cities, increasig from 1,049 units in 1970 to 1,457 in 2010. The greatest concentration of population is in Suffolk, which includes Boston. Furthermore, after 1970, heterogeneity in terms of population size across counties appears jointly with county variance heteroscedasticity.[7]

Among the social, economic, and geographic factors that may affect the population density distribution, a key variable in our analysis is the distance of each census tract from Boston (previously d_{it} and henceforth *Distance*). Because our unit of reference - census tracts- changes over time, *Distance* values are also time-variant and adapt to territorial changes.

The degree of spatial autocorrelation in our data has been measured by Moran's I index for the population log-densities per year. The statistics in Table 1 show a quite strong positive global spatial

	1970	1980	1990	2000	2010
Moran's I	0.76	0.78	0.78	0.80	0.82

Table 1: Moran's I-index for census tract population log-density in Massachussets.

dependence that has slightly increased over time.

For features to represent neighborhood status, we follow Epifani and Nicolini (2017) in using ethnic composition along with education and income as potential quantitative determinants of the population density distribution.

For the ethnic composition of the population, we use the proportion of whites over the total population (henceforth *Mix*). As discussed in Quigley (1985), people are usually more prone to positive discrimination among citizens belonging to the same ethnic group, hence the clustering by ethnic group.

Additionally, in Topa and Zenou's (2015) study on social networks, individuals usually prefer to settle close to other individuals who share the same level of income or education.

For income, we introduce two local predictors: the average income per capita of the previous year (henceforth *Income*)[8] and the Gini index of the income distribution (henceforth *Gini*), which measures income dispersion and inequality. The US census provides the income distribution for each census tract in four classes: less than \$10 000, \$10 000 - \$15 000, \$15 000 - \$25 000, and more than \$25 000.

[7]Refer to Table 1 as well as Figures 1 and 2 in Appendix A.

[8]The variable *Income* is not collected in 1970, so we do not use it in the estimations for that decade.

We computed their Gini index by the formula

$$Gini = \frac{\sum_{k=1}^{n-1}(P_k - Q_k)}{\sum_{k=1}^{n-1} P_k}$$

where Q_k's are the actual cumulative frequencies and P_k's are the cumulative frequencies of income if it was equally distributed.[9]

By contrast, to have a comprehensive indicator of the distribution of the level of education in each census tract, we elaborated a synthetic measure of the degree of education (henceforth *Education*) by ranking all the census tracts according to the level of education of their residents from age 25 onward. Since US census data provide the distribution of education between primary, college, and high levels, we first ranked the census tracts (a) according to the relative frequency of citizens with a high degree of education and, separately, (b) according to the relative frequency of citizens with primary education. Second, for each census tract, we subtracted rank (b) from rank (a). This type of index represents the extent to which a census tract may emerge as highly educated with respect to the rest of the census tracts in Massachusetts.

Finally, some geographical features of the counties have been included in the study by means of the amenities and the surface of the counties. The presence of amenities in each county g is proxied by its proportion Z_g of water areas. Glaeser and Ward (2009) argue that water can be considered a fundamental factor in creating recreational spaces for leisure time. Following Epifani and Nicolini (2017), Z_g can be considered to be constant over time. Regarding the spatial surface S_g of each county g, Combes and Gobillon (2015) suggest that the land (or surface) area is fundamental for picturing the size of agglomeration effects. They conclude that agglomeration gains can stem both from density and from the physical extent of a spatial unit. More precisely, when holding population density constant, the impact of changes in the area (of a spatial unit) reflects agglomeration gains. Therefore, they conclude that agglomeration gains exist when the estimates for land are positive.

The list of our covariates is completed by the interaction between *Distance* and *Income*. It usually amplifies the attractiveness of each tract unit, thereby emphasizing the presence of a mass effect for a number of selected features -in our case, income. The purpose of using that term is to capture the attractiveness of a destination by focusing on some privileged aspects that shape individual preferences for location choices despite the physical distance from the CBD. Put differently, the interaction effects exhibit a spatial autoregressive structure (Wang and Kockelman, 2009).

[9] The Gini index varies between 0 and 1; it is equal to 0 when there is no inequality and reaches 1 when an individual earns the entire income.

In light of the arguments put forward previously and the (statistically significant) spatial dependence in the data structure, we elaborate a step-by-step strategy of analysis. Our first concern is to define the proper empirical framework to perform the econometric analysis, keeping in mind the theoretical setting from Section 3. In the next section, we devote our interest to selecting the functional form with the best fit to accommodate the analysis of the spatial structure and neighborhood status as factors that impact individuals' location decisions. Once we provide this piece of evidence, we turn to refining the way local spatial dependence can be included in the framework of the analysis in order to capture the potential spillovers effects across spatial units (here, census tracts). We perform this second part of the analysis in Section 6, where our preferred device will be the modeling of the degree of spatial dependence by means of census and county random effects according to a few alternative specifications combined in seven different models. The model with the best fit will be retained for drawing corresponding economic insights.

5 Hierarchical Bayesian spatio-temporal Lognormal Model

The definition of a suitable empirical model to perform our analysis involves the building of a framework of analysis that begins with a baseline model including the neighborhood status features (that is the level of education, ethnic composition, income, and amenities) as exogenous predictors and a spatial structure (that is, the physical distance to Boston), but without spatial dependence. Then, this setting will be augmented with features for embedding the local spatial dependence across spatial units.

The baseline lognormal likelihood for the density of the population Y_{it} of census tract i at decade $t = 1, 2, \ldots, T$,[10] henceforth *Model 1*, takes the form

$$\log Y_{it} \,|\gamma_{0t}, \boldsymbol{\beta}_t, \sigma^2, \boldsymbol{\nu} \overset{\text{indep.}}{\sim} \mathcal{N}\Big(\gamma_{0t} + \mathbf{X}^{(1)}_{g(i)}\boldsymbol{\beta}^{(1)}_t + \mathbf{X}^{(2)}_{it}\boldsymbol{\beta}^{(2)}_t, \sigma^2 \times \nu_{g(i)}\Big) \tag{1}$$

where γ_{0t} is a time-decade intercept to be estimated, the index function $g(i)$ returns the county g to which census tract i belongs, and $\mathbf{X}^{(1)}_g = [Z_g, S_g]$ is the row vector of amenities Z_g and surface S_g; Z_g, S_g are constant over time and associated with unknown coefficients $\boldsymbol{\beta}^{(1)}_t$. In addition, $\mathbf{X}^{(2)}_{it}$ is the row vector containing the time-variant census-level covariates *Distance*, *Mix*, *Education*, *Gini*, and *Income* as well as the interaction *Distance*$\times$*Income*, associated with the unknown coefficients

[10]The definition $t = 1$ refers to the first decade 1970, 2 refers to the second decade 1980, and $\ldots$, T refers to the fifth decade 2010.

$\boldsymbol{\beta}_t^{(2)}$.[11] To manage tractable values, all the exogenous predictors in $[\mathbf{X}_g^{(1)}, \mathbf{X}_{it}^{(2)}]$ were standardized by subtracting their sample mean and normalizing by their sample standard deviation. For the sake of simplicity, Table 2 summarizes the dependent variable and all the covariates. The common unknown

Y_t	$\mathbf{X}^{(1)}$		$\mathbf{X}_t^{(2)}$					
Population's Density	Z *Amenities*	S *Surface*	$X_{1t}^{(2)}$ *Distance*	$X_{2t}^{(2)}$ *Mix*	$X_{3t}^{(2)}$ *Education*	$X_{4t}^{(2)}$ *Gini*	$X_{5t}^{(2)}$ *Income*	$X_{6t}^{(2)} := X_{1t}^{(2)} \times X_{5t}^{(2)}$ *Distance×Income*

Table 2: Legend for population's density Y_t and selected covariates.

variance σ^2 aims to capture the implicit variability shared by the population log-density of all census tracts at each time t, and the vector $\boldsymbol{\nu} := (\nu_1, \cdots, \nu_{14})$ measures the county-specific variabilities. Following LeSage and Pace (2009), Equation (1) accommodates the county heteroscedasticity in a Bayesian spirit by the multiplicative structure $\sigma_g^2 = \sigma^2 \times \nu_g$, $g = 1, \ldots, 14$.

According to a hierarchical *specific-to-general* scheme, including the spatial dimension allows for the possibility of augmenting *Model 1* with *i*) exogenous spatial interaction effects among covariates involving neighboring census tracts, *ii*) interaction among random county effects, and *iii*) interaction among census random effects. The spatial model that includes all three types of spatial effects takes the form

$$(2)\quad \log Y_{it} |\, \phi_{it}, \boldsymbol{\gamma}_t, \boldsymbol{\beta}_t, \boldsymbol{\delta}_t, \sigma^2, \boldsymbol{\nu} \overset{\text{indep.}}{\sim} \mathcal{N}\Big(\gamma_{g(i)t} + \phi_{it} + \mathbf{X}_{g(i)}^{(1)} \boldsymbol{\beta}_t^{(1)} + \mathbf{X}_{it}^{(2)} \boldsymbol{\beta}_t^{(2)} + \mathbf{W}_{it} \mathbf{X}_t^{(2)} \boldsymbol{\delta}_t, \sigma^2 \times \nu_{g(i)}\Big).$$

We refer to the model identified by the set of Equations (2)–(5) as *Model 6*. It replicates the distinguishing components of the SDEM - that is the exogenous spatial interaction between the covariates and spatial dependence in the disturbances.

Vector $\mathbf{W}_{it}$ is the ith row of the first-order row-stochastic contiguity matrix $\mathbf{W}_t$ that has null elements on the principal diagonal and, for $i \neq j$: $w_{ijt} = 1/m_{it}$ if census tracts i, j are neighbors at time t (i.e., m_{it} is the number of neighboring census tracts of i at time t) and $w_{ijt} = 0$ otherwise;[12] $\mathbf{W}_{it}\mathbf{X}_t^{(2)}$ is the vector of the spatially lagged explanatory variables. Its introduction proposes that population density can be affected by changes in explanatory variables in the neighboring census tracts summarized by their average. The direct effects at decade t are the estimates of the coefficients $\boldsymbol{\beta}_t^{(1)}, \boldsymbol{\beta}_t^{(2)}$ associated with the variables $\mathbf{X}_g^{(1)}, \mathbf{X}_t^{(2)}$, and the estimate of $\boldsymbol{\delta}_t$ is the vector of the spillovers (or indirect) effects associated with $\mathbf{W}_t\mathbf{X}_t^{(2)}$.

The random vector $\boldsymbol{\gamma}_t = (\gamma_{1\,t}, \ldots, \gamma_{14\,t})$ refers to 14 county random effects at time t. Each frailty

[11] According to the setting in Section 3, the coefficients $\boldsymbol{\beta}_t$ have been split into two disjointed subsets $\boldsymbol{\beta}_t = [\boldsymbol{\beta}_t^{(1)}, \boldsymbol{\beta}_t^{(2)}]$, one referring to covariates at the county level and the other at the census-tract level.

[12] Notice that some neighboring census tracts of i may fall outside census tract i's county.

γ_{gt} captures all the predictors of the population density, whether unobservable or observable, common to all the census tracts in county g, but not measured. Involving all the census tracts in the same county g, frailty γ_{gt} potentially captures a dimension of county effects that are not fully embedded in natural amenities Z_g or in spatial surface S_g. Our rationale is that census tracts belonging to the same county share some common features (e.g., urban regulation) that individuals consider in forming their preferences. Hence, the frailties in $\boldsymbol{\gamma}_t$ model global mean *random* heterogeneity -namely, global spatial dependence. All γ_{gt} frailties are assumed to be conditionally independent with a common constant variance σ^2_γ, given the time-decade intercept γ_{0t} that satisfies

$$(3) \qquad \gamma_{gt} \mid \gamma_{0t}, \sigma^2_\gamma \overset{\text{indep.}}{\sim} \mathcal{N}(\gamma_{0t}, \sigma^2_\gamma).$$

Moreover, the common hyperparameter γ_{0t} is ruled by the autoregressive dynamics

$$(4) \qquad \gamma_{01} \sim \mathcal{N}(0, \sigma^2_0) \quad \text{and} \quad \gamma_{0t} \mid \gamma_{0t-1} \sim \mathcal{N}(\gamma_{0t-1}, \sigma^2_0), \quad t = 2, \ldots, T$$

that yield a correlation among all γ_{gt} across space and time given by

$$\rho(\gamma_{gs}, \gamma_{ht}) = \frac{\min\{s,t\}}{\sqrt{(\min\{s,t\}+a)(\max\{s,t\}+a)}} \quad \text{if } s \neq t \text{ or } g \neq h \quad \text{with} \quad a := \frac{\sigma^2_\gamma}{\sigma^2_0}.$$

The formula for the correlation $\rho(\gamma_{gs}, \gamma_{ht})$ sheds light on the meaning of the ratio a of the variance σ^2_γ to σ^2_0. Indeed, $\rho(\gamma_{gs}, \gamma_{ht})$ is a decreasing function of a, and the larger is σ^2_γ with respect to σ^2_0, the closer the correlation $\rho(\gamma_{gs}, \gamma_{ht})$ is to zero. Equations (3)-(4) embed the same spatial dependence across any two counties at time t but changing over time. Briefly, the population density in one county at time t depends on the county's population density at time s previous to t as well as on the current and past population density of the other counties; for any fixed s, the correlation $\rho(\gamma_{gs}, \gamma_{ht})$ approaches zero as t approaches infinity. This way of modeling is flexible enough for obtaining a county-specific correlation: it is sufficient to assume different variances $\sigma^2_{\gamma_g}$ for the random effects γ_{gt} in Equations (3)-(4).[13] In this sense, $\rho(\gamma_{gs}, \gamma_{ht})$ serves as an index of association between counties.

Introducing spatial CAR census random effects is the principal technical novelty of our contribution. The T vectors of the census random effects $(\phi_{11}, \ldots \phi_{n_1 1}), \ldots, (\phi_{1T}, \ldots \phi_{n_T T})$ allow for controlling for the random mean heterogeneity at the census tract level (hence, the local spillovers). They are time-independent, and each of them has been modeled as a *proper* CAR random process, meaning

[13]The econometric analysis of county spatial models developed by Epifani and Nicolini (2018) shows that no sizable difference exists between adopting a CAR strategy for county frailties or modeling them with Equations (3)-(4).

that at each decade t, the full conditional distribution of the vector $(\phi_{1t}, \dots \phi_{n_t t})$ is

$$\phi_{it} \mid \phi_{jt},\ j \neq i, \alpha_t, \tau_t \ \sim \mathcal{N}\Big(\alpha_t \sum_{j \neq i} w_{ijt}\phi_{jt}\ ,\ \frac{\tau_t}{m_{it}}\Big). \tag{5}$$

Since w_{ijt} is the (ij) element of the row-stochastic contiguity matrix $\mathbf{W}_t$ of order 1, then $\sum_{j \neq i} w_{ijt}\phi_{jt}$ can be interpreted to be the arithmetic mean of the ϕ_{jt}'s neighboring of unit i at time t. Therefore, the census random effect ϕ_{it} is centered on the arithmetic mean of its neighbors. The strength of the spatial dependence between the nearest neighboring census tracts is quantified by the unknown spatial autoregressive parameter α_t. By Brook's lemma, the joint distribution of $(\phi_{1t}, \dots \phi_{n_t t})$ is multivariate normal with

$$(\phi_{1t}, \dots \phi_{n_t t}) \sim \mathrm{MN}(\mathbf{0}, \tau_t[\mathbf{D}_t - \alpha_t \mathbf{W}_t]^{-1}),$$

where $\mathbf{D}_t = diag(m_{it})$ is a diagonal matrix; if α_t belongs to $[0, 1)$, then the covariance matrix $\tau_t[\mathbf{D}_t - \alpha_t \mathbf{W}_t]^{-1}$ is positive definite and the multivariate normal distribution $\mathrm{MN}(\mathbf{0}, \tau_t[\mathbf{D}_t - \alpha_t \mathbf{W}_t]^{-1})$ is proper. If $\alpha_t = 1$, the proper CAR model in (5) collapses to an *intrinsic* CAR model that lacks a probability density function. Instead, the parameter value $\alpha_t = 0$ corresponds to time-independent census tracts:

$$\phi_{it} \mid \sigma^2_{\phi,t} \overset{\text{indep.}}{\sim} \mathcal{N}\left(0, \frac{\tau_t}{m_{it}}\right) \quad \forall i, t. \tag{6}$$

For that reason, α_t is defined a priori in the interval $[0, 1)$; a large value of α_t confirms the existence of a strong spatial correlation.[14]

Before moving to estimations, in a Bayesian framework, we need to specify the prior distribution scheme for the unknown parameters $\boldsymbol{\beta}_t^{(1)}, \boldsymbol{\beta}_t^{(2)}, \boldsymbol{\delta}_t, \sigma^2, \boldsymbol{\nu}, \sigma^2_0,\ \sigma^2_\gamma, \tau_t, \alpha_t$.

A priori, we assume an autoregressive hierarchical model for the sequences of the intercept $(\gamma_{0t})_t$, the direct effects $(\boldsymbol{\beta}_t)_t$, and the spillovers $(\boldsymbol{\delta}_t)_t$. In particular, the block of the regression parameters $\boldsymbol{B}_t := (\gamma_{0t}, \boldsymbol{\beta}'_t, \boldsymbol{\delta}'_t)$, $t = 1, \dots, T$ has the prior dynamics given by

$$\begin{cases} \boldsymbol{B}_1 \sim \mathrm{MN}(\mathbf{0}, \Sigma_{\boldsymbol{B}}) \\ \boldsymbol{B}_t | \boldsymbol{B}_{t-1} \sim \mathrm{MN}(\boldsymbol{B}_{t-1}, \Sigma_{\boldsymbol{B}}) \quad \text{for } t = 2, \dots, T \\ \Sigma_{\boldsymbol{B}} = diag\{\sigma^2_0, \sigma^2_{\beta^{(1)}_1}, \sigma^2_{\beta^{(1)}_2}, \sigma^2_{\beta^{(2)}_1}, \dots, \sigma^2_{\beta^{(2)}_6}, \sigma^2_{\delta_1}, \dots, \sigma^2_{\delta_6}\}, \end{cases}$$

[14] For more details on CAR models, see Banerjee et al. (2015).

where the variance hyperparameters $\sigma^2_0, \sigma^2_{\beta^{(1)}_1}, \sigma^2_{\beta^{(1)}_2}, \sigma^2_{\beta^{(2)}_1}, \ldots, \sigma^2_{\beta^{(2)}_6}, \sigma^2_{\delta_1}, \ldots, \sigma^2_{\delta_6}$ are modeled a priori in terms of their standard deviations in the usual way to be diffuse:

$$\sigma_0, \sigma_{\beta^{(1)}_1}, \sigma_{\beta^{(1)}_2}, \sigma_{\beta^{(2)}_1}, \ldots, \sigma_{\beta^{(2)}_6}, \sigma_{\delta_1}, \ldots, \sigma_{\delta_6} \overset{iid}{\sim} \text{Uniform}(0, 100).$$

Taken from the theory of dynamic linear models introduced by West and Harrison (1997), the random walk prior specified for the direct and indirect effects $\boldsymbol{\beta}_t, \boldsymbol{\delta}_t$ embeds the time dimension. This prior is quite comparable to the canonical way of dealing with the classic idea of adaptive expectations, meaning that we are introducing a sort of memory in the dynamics process in an autoregressive style.

The prior variances' structure of the population log-densities σ^2_g, $g = 1, \ldots, 14$, is

$$\begin{cases} \sigma^2_g = \sigma^2 \times \nu_g, \quad \sigma^2, \nu_1, \nu_2, \ldots, \nu_{14} \quad \text{independent} \\ \sigma^2 \sim \text{InverseGamma}(0.001, 0.001) \\ \nu_1, \nu_2, \ldots, \nu_{14} \overset{iid}{\sim} \text{InverseGamma}\Big(\frac{r}{2}, \frac{r}{2}\Big), \end{cases} \tag{7}$$

where InverseGamma$(r/2, r/2)$ stands for the inverse gamma distribution with shape and rate $r/2$, and $\sigma^2, \boldsymbol{\nu}$ are independent of all the other parameters. If a large value were assigned to parameter r, then all of the ν_g's a priori would collapse into one and the assumption of constant variance over the counties would hold. Conversely, small values of r cause a skewed distribution of ν_g's, which allows for both large and small values of σ^2_g and entails a strong heteroscedasticity among counties. In this sense, formulae in (7), from LeSage and Pace (2009), accommodate Bayesian heteroscedasticity in spatial models. Here, we define $r = 4$ as suggested by LeSage and Pace (2009).

A priori the standard deviation σ_γ, common to all county random effects in $\boldsymbol{\gamma}_t$, comes from the diffuse uniform density on $(0, 100)$.

Last, we assume prior independence of the hyperparameters α_t, τ_t of the CAR census effects $(\phi_{1t}, \ldots \phi_{n_t t})$ for $t = 1, \ldots, T$, such that

$$\text{logit}(\alpha_t) \overset{iid}{\sim} \text{N}(0, 0.02), \quad \tau_t \overset{iid}{\sim} \text{InverseGamma}(0.5, 0.0005) \qquad \forall t = 1, \ldots, T.$$

Variances τ_t's change a priori over time since census tracts cannot be mapped over time; therefore, in each decade, we handle a different sample.[15]

[15] See Appendix B for a discussion of the prior choice for α_t, τ_t.

6 Goodness-of-fit and model selection

The previous lognormal model allows for formalizing six different empirical forms that we label *Model 1–Model 6*. The steps for moving from *Model 1* to *Model 6* involve defining a series of nested models that can be formalized by adding spatial interaction effects to *Model 1* at different levels. *Model 2* embeds only the exogenous interaction effect $\mathbf{W}_{it}\mathbf{X}_t^{(2)}$ among covariates. *Model 2* is an SLX model, which is considered the simplest econometric model that is able to flexibly accommodate the study of spatial spillovers (Halleck Vega and Elhorst, 2015). Its structure can be enriched by introducing either county random frailties $\boldsymbol{\gamma}_t$ to better control for global spatial dependence (*Model 3*) or, alternatively, independent census random effects $(\phi_{1t}, \ldots \phi_{n_t t})$ (*Model 4*) to capture local spatial heterogeneity, or introducing both (*Model 5*). Along this line, *Model 6* generalizes *Model 5* by imposing a CAR structure on census effects $(\phi_{1t}, \ldots \phi_{n_t t})$. With this in mind, we have to identify the most effective framework for modeling spatial dependence.

We assessed the goodness-of-fit of our models by computing the Bayesian posterior p-value p_M of each *Model M* (for $M = 1, \ldots, 6$) on the basis of the χ^2 discrepancy measure between the data of population density and their prediction under *Model M*, as suggested in Gelman et al. (1996). A value for p_M of around 0.5 indicates that *Model M* fits quite well, whereas values close to zero or one suggest a limited fitting capacity. The estimated p_M are summarized in the first column of Table 3.

A Bayesian comparison of *Models 1-6* was performed by computing the value of the Bayesian Deviance Information Criterion (DIC) of every model, the Logarithm of the Pseudo-Marginal Likelihood ($LPML$), and the Bayesian percentage census-tract outliers with level 90% of every model at each decade. Models with a small DIC value are preferred to those with a large DIC value. Instead, the larger the value of the $LPML$ at time t, the better the fit of the model at that time. The DIC values appear in the second column of Table 3, and the 90% Bayesian percentage outliers and $LPML$ statistics per decade are presented in the last five columns of Table 3. The computation of $p_M, DIC, LPML$, and the Percentage of the Bayesian census-tract outliers has been performed via Markov Chain Monte Carlo simulation (see Appendix C for more technical details).

Overall, our specifications perform relatively well since they produce relatively good p_M at approximately 0.5. The percentage of Bayesian outliers in each of the six models remains quite low (less than 10%) for the period from 1980 to 2010, whereas values for 1970 are at odds with the rest of the sample because of the political and economic changes experienced in Massachusetts and discussed in Epifani and Nicolini (2013). DIC and $LPML$ values suggest that the best specification has to embed

	p_M	DIC		1970	1980	1990	2000	2010
Model 1	0.505	19474.5	Bayes Outliers	24.31%	8.5%	7.22%	5.01%	4.32%
			$LPML$	−1960.360	−1685.272	−1860.977	−1831.775	−1941.548
Model 2	0.494	18590.1	Bayes Outliers	24.02%	8.01%	5.45%	4.86%	4.39%
			$LPML$	−1963.865	−1586.516	−1729.097	−1720.948	−1839.882
Model 3	0.506	17858.6	Bayes Outliers	23.83%	8.53%	5.45%	4.79%	4.12%
			$LPML$	−1871.515	−1536.946	−1656.540	−1644.998	−1767.934
Model 4	0.505	17655.8	Bayes Outliers	0.10%	6.82%	6.76%	7.52%	7.41%
			$LPML$	−1737.717	−1582.664	−1707.496	−1688.804	−1798.088
Model 5	0.518	16971.9	Bayes Outliers	0.00%	4.94%	5.15 %	8.18%	8.03%
			$LPML$	−1697.888	−1542.292	−1641.051	−1609.181	−1724.413
Model 6	0.505	5067.8	Bayes Outliers	0.00%	0.26%	0.46 %	0.44 %	0.41%
			$LPML$	−645.402	−698.225	−805.926	−795.080	−828.763

Table 3: Bayesian posterior p_M-values and DIC of *Models 1-6* in the first two columns; Percentage of Bayesian 90% Outliers (first row) and $LPML$ (second row) per decade in the last five columns.

spatial CAR dependence between the census tracts. In that regard, *Model 6* clearly outperforms the others with respect to all the indicators.

A second step of the model selection analysis focuses on the posterior estimate of county random heterogeneity. Our strategy consists of assessing the role of county random effects γ_{gt}'s when introducing a CAR census effect, for we want to test whether γ_{gt}'s (thus, global spatial effects) retain statistical significance in shaping the population density distribution even in the presence of local spatial effects. To that end, we computed the posterior median values of γ_{gt}'s for each county g and decade t in *Models 3, 5, and 6* that include them. Results are illustrated in Figure 1, which shows that imposing a CAR structure into the census random effects, as in *Model 6*, dampens the impact of the hierarchical county structure, especially in the most recent decades. These outcomes imply a clustering effect at the county level that captures a somewhat spatial-dependent structure in the absence of spatially correlated census random effects, as in *Models 3* and *5*. But when introducing CAR census effects, a part of that county effect is captured by the CAR structure. Figure 1 emphasizes that this CAR effect is more prominent when referring to the counties with census tracts that are smaller in size (namely, the ones in Suffolk, Worcester, and Middlesex) and for which connectivity and network interaction across neighborhood units are more likely. Indeed, the posterior medians of these county effects approach zero. In this respect, our CAR structure turns out to be effective in capturing local spatial dependence, as is usually intended in the literature. This is reinforced by the effective local spatial dependence that exists across neighboring census tracts, as clarified in the comparison between *Model 5* with all a priori independent census frailties and *Model 6* with CAR census frailties.[16]

The previous findings suggest exploring the goodness-of-fit of a simplified version of *Model 6* that

[16]Estimations for direct and indirect effects (see Section 7) are robust irrespective of the presence of the county random effects too. Results are available upon request.

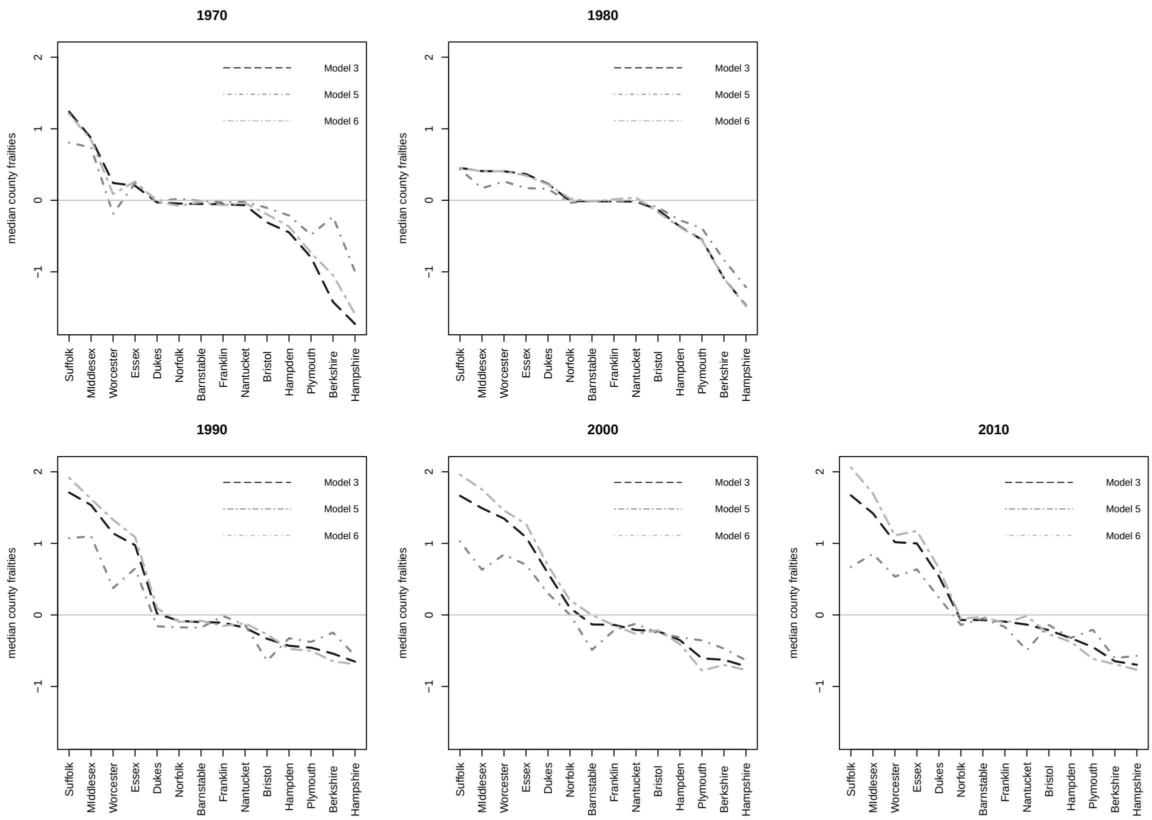


Figure 1: Posterior medians of the county frailties $\boldsymbol{\gamma}_t$ under *Models 3;5;6*. Data are not available for census tracts in Barnstable, Franklin, Hampden and Norfolk in 1970, 1980. Counties are sorted horizontally in decreasing order according to the frailty posterior medians estimated under the simplest *Model 3*.

includes spatially lagged covariates and CAR census effects only:

$$(8) \quad \log Y_{it} | \gamma_{0t}, \phi_{it}, \boldsymbol{\beta}_t, \boldsymbol{\delta}_t, \sigma^2, \boldsymbol{\nu} \overset{\text{indep.}}{\sim} \mathcal{N}\Big(\gamma_{0t} + \phi_{it} + \mathbf{X}^{(1)}_{g(i)} \boldsymbol{\beta}^{(1)}_t + \mathbf{X}^{(2)}_{it} \boldsymbol{\beta}^{(2)}_t + \mathbf{W}_{it} \mathbf{X}^{(2)}_t \boldsymbol{\delta}_t, \sigma^2 \times \nu_{g(i)}\Big)$$

Our final *Model 7* is specified by the likelihood (8) and the CAR modeling of ϕ_{it}'s in (5). It emphasizes local spatial dependence exclusively; thus, it is a SDEM-type model with CAR disturbances ϕ_{it}'s. After performing the canonical Bayesian goodness-of-fit and model selection analysis, we show in

	p_M	DIC		1970	1980	1990	2000	2010
Model 7	0.496	**4077.0**	Bayes Outliers	0%	0.17%	0.31%	0.22%	0.41%
			LPML	**−553.445**	**−615.534**	**−739.822**	**−702.266**	**−744.244**

Table 4: Bayesian posterior p_M-value, DIC, Percentage of Bayesian 90% Outliers, and *LPML* per decade for *Model 7*.

Tables 3 and 4 that *Models 6* and *7* are definitely more efficient than the other competing models.

But according to all of these indicators, *Model 7* turns out to be the most efficient specification.

As a further indicator for the assessment of the most suitable model for our analysis, given the importance of CAR effects, an interesting (and more sophisticated) statistical indicator to take into account for the model choice is the pattern of CAR effects across time. Table 5 presents the posterior estimate of the percentage of non-null census random effects under *Models 4–7*. *Model 7* records

	1970	1980	1990	2000	2010
Model 4	17.06%	0.00%	0.00%	0.00%	0.00%
Model 5	12.68%	0.26%	0.00%	0.00%	0.00%
Model 6	62.82%	49.19%	47.31%	45.91%	47.29%
Model 7	70.45%	58.06%	55.38%	55.93%	54.84%

Table 5: Percentage of posterior non-null census random effects in relation to the corresponding 95% credible interval for each model specification including either independent or CAR ϕ_{it}'s.

more than 50% of significant (i.e., non-null) census random effects, whereas that percentage drops in the case of independent census random effects. Such statistics confirm that census random effects are effective in modeling the population density distribution at any time, although their effectiveness improves when a CAR structure is introduced. Given that evidence at hand, the key role for census tracts might be a consequence of improvements in the transportation system - for instance, changes that improve the accessibility to different spatial units and favor the redistribution of agents across a territory.[17]

7 Bayesian estimation: direct and indirect effects

According to goodness-of-fit and model selection analysis, *Model 7* is the most effective framework to estimate the determinants that shape the population density distribution.

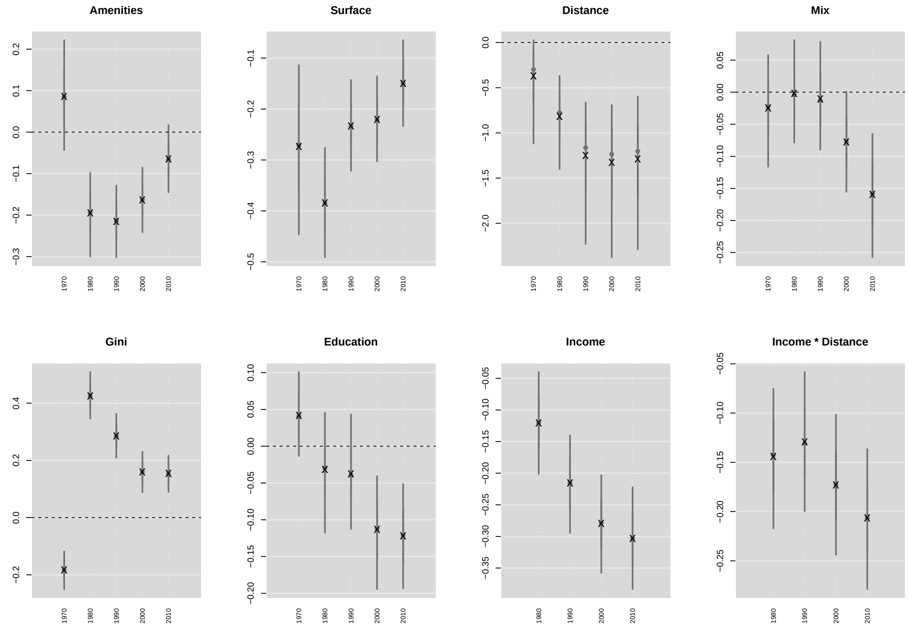
Preliminary descriptive statistics reveal the spatial dimension to be crucial, and our Bayesian estimations embed it. Estimates for $\boldsymbol{\beta}_t^{(1)}, \boldsymbol{\beta}_t^{(2)}, \boldsymbol{\delta}_t$ appear in Table 6 and Figure 2. Furthermore, given that the final structure of *Model 7* replicates a SDEM, we can easily discuss direct and indirect effects. Importantly, the spatial autocorrelation among census tracts appears positive and important, as documented in the last row of Table 6, which contains a summary of the posterior distribution for the spatial autoregressive coefficients α_t, $t = 1, \ldots, 5$.

[17]This result aligns the content of Figure 5, which depicts the estimated temporal evolution of CAR census random frailties versus the distance from Boston, as well as the maps in Figures 3-7 in Appendix D, which display the evolution of the CAR census random effects in Massachusetts, under *Model 7*.

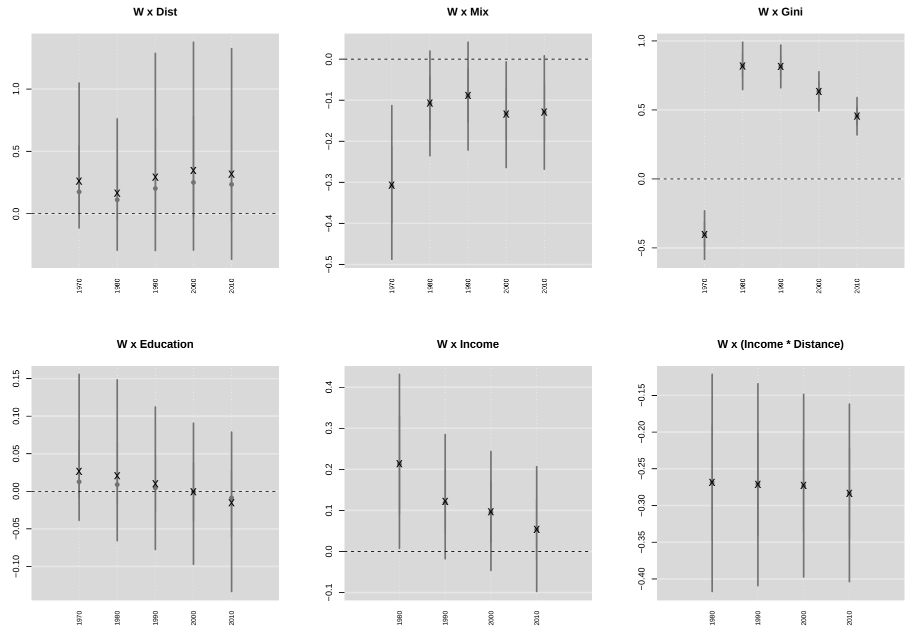
Response Parameters (Covariates)	1970	1980	1990	2000	2010
Intercept γ_{0t} for time t	0.217 (0.079) $(0.068, 0.378)$ 0.002	0.146 (0.054) $(0.039, 0.250)$ 0.008	0.010 (0.048) $(-0.086, 0.104)$ 0.829	−0.066 (0.045) $(-0.154, 0.021)$ 0.137	−0.120 (0.047) $(-0.212, -0.031)$ 0.009
$\beta_1^{(1)}$ (Amenities)	0.086 (0.069) $(-0.043, 0.222)$ 0.210	−0.195 (0.051) $(-0.300, -0.098)$ 0.000	−0.215 (0.044) $(-0.302, -0.129)$ 0.000	−0.163 (0.040) $(-0.241, -0.085)$ 0.000	−0.065 (0.042) $(-0.145, 0.017)$ 0.122
$\beta_2^{(1)}$ (Surface)	−0.274 (0.084) $(-0.446, -0.114)$ 0.001	−0.384 (0.055) $(-0.491, -0.276)$ 0.000	−0.233 (0.046) $(-0.322, -0.143)$ 0.000	−0.221 (0.042) $(-0.303, -0.136)$ 0.000	−0.149 (0.043) $(-0.234, -0.065)$ 0.000
$\beta_1^{(2)}$ (Distance from CBD)	−0.371 (0.304) $(-1.118, 0.026)$ 0.038	−0.819 (0.261) $(-1.400, -0.370)$ 0.001	−1.251 (0.400) $(-2.229, -0.662)$ 0.000	−1.328 (0.436) $(-2.376, -0.692)$ 0.000	−1.286 (0.431) $(-2.288, -0.599)$ 0.000
$\beta_2^{(2)}$ (Mix)	−0.025 (0.044) $(-0.116, 0.058)$ 0.284	−0.002 (0.040) $(-0.079, 0.081)$ 0.457	−0.011 (0.043) $(-0.090, 0.078)$ 0.379	−0.077 (0.039) $(-0.155, 0.001)$ 0.026	−0.159 (0.050) $(-0.257, -0.065)$ 0.001
$\beta_3^{(2)}$ (Gini)	−0.184 (0.033) $(-0.250, -0.118)$ 0.000	0.426 (0.042) $(0.346, 0.509)$ 0.000	0.285 (0.039) $(0.209, 0.363)$ 0.000	0.160 (0.035) $(0.088, 0.230)$ 0.000	0.153 (0.032) $(0.090, 0.216)$ 0.001
$\beta_4^{(2)}$ (Education)	0.042 (0.030) $(-0.013, 0.021)$ 0.078	−0.031 (0.041) $(-0.118, -0.056)$ 0.213	−0.038 (0.039) $(-0.112, -0.064)$ 0.160	−0.113 (0.039) $(-0.194, -0.138)$ 0.001	−0.122 (0.037) $(-0.193, -0.147)$ 0.000
$\beta_5^{(2)}$ (Income)	NA NA NA	−0.121 (0.041) $(-0.201, -0.040)$ 0.002	−0.216 (0.039) $(-0.294, -0.140)$ 0.000	−0.279 (0.040) $(-0.358, -0.203)$ 0.000	−0.303 (0.041) $(-0.383, -0.222)$ 0.000
$\beta_6^{(2)}$ (Income*Distance)	NA NA NA	−0.144 (0.037) $(-0.217, -0.075)$ 0.000	−0.129 (0.036) $(-0.200, -0.058)$ 0.000	−0.173 (0.035) $(-0.244, -0.102)$ 0.000	−0.207 (0.036) $(-0.279, -0.137)$ 0.000
δ_1 ($W \times$ Distance)	0.260 (0.312) $(-0.114, 1.049)$ 0.167	0.164 (0.266) $(-0.293, 0.761)$ 0.271	0.294 (0.401) $(-0.295, 1.287)$ 0.225	0.345 (0.435) $(-0.291, 1.376)$ 0.213	0.319 (0.434) $(-0.366, 1.324)$ 0.232
δ_2 ($W \times$ Mix)	−0.306 (0.095) $(-0.488, -0.371)$ 0.000	−0.107 (0.065) $(-0.235, -0.149)$ 0.049	−0.089 (0.067) $(-0.221, -0.134)$ 0.093	−0.133 (0.065) $(-0.264, -0.176)$ 0.020	−0.129 (0.070) $(-0.268, -0.176)$ 0.031
δ_3 ($W \times$ Gini)	−0.403 (0.089) $(-0.583, -0.462)$ 0.000	0.819 (0.088) $(0.648, 0.760)$ 0.000	0.814 (0.079) $(0.660, 0.760)$ 0.000	0.634 (0.073) $(0.492, 0.583)$ 0.000	0.455 (0.068) $(0.319, 0.409)$ 0.000
δ_4 ($W \times$ Education)	0.026 (0.048) $(-0.039, -0.002)$ 0.295	0.020 (0.051) $(-0.066, -0.007)$ 0.361	0.010 (0.045) $(-0.078, -0.014)$ 0.427	−0.001 (0.045) $(-0.097, -0.023)$ 0.490	−0.016 (0.051) $(-0.133, -0.043)$ 0.394
δ_5 ($W \times$ Income)	NA NA NA	0.213 (0.113) $(0.008, 0.431)$ 0.015	0.122 (0.078) $(-0.018, 0.285)$ 0.049	0.096 (0.075) $(-0.046, 0.244)$ 0.090	0.054 (0.076) $(-0.097, 0.207)$ 0.230
δ_6 ($W \times$ Income*Distance)	NA NA NA	−0.269 (0.077) $(-0.417, -0.121)$ 0.000	−0.271 (0.070) $(-0.409, -0.134)$ 0.000	−0.272 (0.064) $(-0.397, -0.149)$ 0.000	−0.283 (0.062) $(-0.404, -0.162)$ 0.000
α_t (spatial CAR parameter)	0.842 (0.014) $(0.813, 0.868)$	0.796 (0.018) $(0.759, 0.830)$	0.763 (0.021) $(0.721, 0.801)$	0.778 (0.019) $(0.739, 0.814)$	0.811 (0.017) $(0.777, 0.843)$

Table 6: Summaries of the posterior distribution of the direct $\boldsymbol{\beta}_t^{(1)}, \boldsymbol{\beta}_t^{(2)}$ and indirect $\boldsymbol{\delta}_t$ effects and the spatial autocorrelation α_t for *Model 7*. Posterior means are followed by (standard deviations) in first row; 95% credible intervals are in second row, Bayesian posterior-predictive p-values are in third row.

Concerning the direct effects, the distance to Boston is the key determinant event if its size shrinks over time. In economic terms, this result aligns with the idea that residents prioritize proximity to services and the urban environment offered by Boston once they discount the benefits of living in a specific county. In turn, the result confirms the existence of competition effects to settle near Boston and the drawback of commuting costs to reach the city (cf. Glaeser, 2008). In that sense, a specific centripetal force seems to exist in Boston that is not likely to be compensated by other types of ag-



(a)



(b)

Figure 2: Plot of the Bayesian posterior means (solid black circles) and 95% credible intervals (gray vertical bars) of the regression coefficients in *Model 7*, from 1970 to 2010 (data in 1970 for *Income* are not available).

glomeration effects at a local level in other spatial units. This type of effect is strongly associated with a single spatial unit, and there are no (statistically significant) indirect effects arising from spillovers from neighborhood units.

As for the spatial dimension, the posterior mean of the coefficient $\beta_{2t}^{(1)}$ associated with the county size S_g is negative, as is the posterior mean of $\beta_{1t}^{(1)}$ associated with the county amenities Z_g; with the exception of 1970, amenities and agglomeration effects at the county level seem to be important in individuals' location choices.

Regarding the other covariates, it is interesting to remark the interplay between distance and other covariates across time. The estimated direct effects suggest that when the size of the coefficient associated with the distance from Boston shrinks, the population density distribution in Massachusetts is enforced by a spread guided by discrimination aptitude. The direct effects associated with the variables *Mix*, *Education*, and *Income* have been mostly statistically significant in recent decades and display negative coefficients. At the same time, the direct effect of the *Gini* index is always positive (excluding 1970). These dynamics suggest a dual reading of the effects. First, (white) residents seem to favor settling in census tracts with a high density of white residents, populated by relatively well-educated people, and in wealthy areas where inhabitants prefer to live in individual dwellings. This dimension of segregation is reinforced by the posterior distribution of the coefficients $\beta_{6t}^{(2)}$ associated with the interaction *Distance*×*Income* that are concentrated in negative intervals. The proposed discrimination-guided attitude in individuals' location decisions is confirmed by the estimate of the *Gini* direct effect as well as by the estimated indirect effects for the other covariates (when statistically significant). A joint reading of all the previous results yields a simple interpretation. Overall, a general attitude driven by discrimination in location choices has been reinforced over time with a positive dependence across all census tracts in a sort of complementarity effect. As already discussed in Epifani and Nicolini (2017), people with high income tend to settle in selective and exclusive locations where they have access to individual dwellings and, hence, where the population density is lower. As a consequence, the concentration of people with similar income levels in a spatial unit reduces the degree of heterogeneity in terms of average income in that unit and, hence, the Gini index approaches zero and the association with a positive (estimated) direct effect yields low population density. It is interesting to note that the results of indirect effects confirm the existence of local spatial dependence, given the existence of a spatial segregation aptitude (mostly driven by the income component). Such behavior is not limited to a spatial unit but spills over into the surrounding units likely for networking mechanisms across individuals, for instance. Hence, the spatial segregation configuration spans over more than a single census tract. However, the local spatial dependence effect is a relatively recent phenomenon. When we focus on the indirect effect for *Income*, we appreciate that the estimated coefficient was positive and statistically significant up the 2000s, albeit with a decreasing magnitude. This result implies that up to that time period, the income effect was a factor driving individual location choices, but without the local spatial dependence component. This condition, for instance, implies that high-income census tracts featuring low density population values could neighbor low-income census tracts with high population density values. This type of spatial organization of land stems from the adoption of a specific regulation for the real estate market that prevents free access to all the residential options for everyone. In this respect, Epifani and Nicolini (2017) discuss important features that limited the development of the housing market and free residential access in the US up to 1980s: limited access to credit or racial discrimination drove groups of citizens to concentrate in selected places. Under this perspective, the highly regulated housing and credit markets limited the potentially positive spillovers effects across census tracts. Once this regulation was repealed and free accessibility to the real estate market was restored, the local spatial spillovers effects became effective in shaping the population distribution. This transforming effect is taken into account in our estimation by the CAR effects that became truly relevant components for the population density distribution from the 1990s onward.

On the whole, the introduction of CAR census effects helps to shape spatial features at the territorial level, and thus improves the model's fit. In Figure 3, which plots the posterior 95% credible intervals and actual values of the population log-density versus *Distance*, our specification captures the highest population densities fairly well.

Regarding heteroscedasticity, Figure 4 shows that the highest variances occur in the most remote counties (i.e., Nantucket, Berkshire, and Hampshire) and are similar, whereas the other counties can be clustered in one group with a smaller population log-density variance. Those results suggest that using CAR random effects can better pinpoint residual heteroscedasticity. Referring to the temporal dimension (and returning to the discussion about free accessibility to the real estate market), Figure 5 reveals the existence of a persistent structural break between the first two decades (1970 and 1980) and the others (1990, 2000, and 2010). In the second period, the core counties (i.e, those closest to Boston) display strictly negative estimated posterior medians of the CAR census effects; the remaining counties (i.e, the ones farther from Boston) either tend toward zero or are strictly positive. Those outcomes stress that people who settled in the most remote counties farthest from Boston

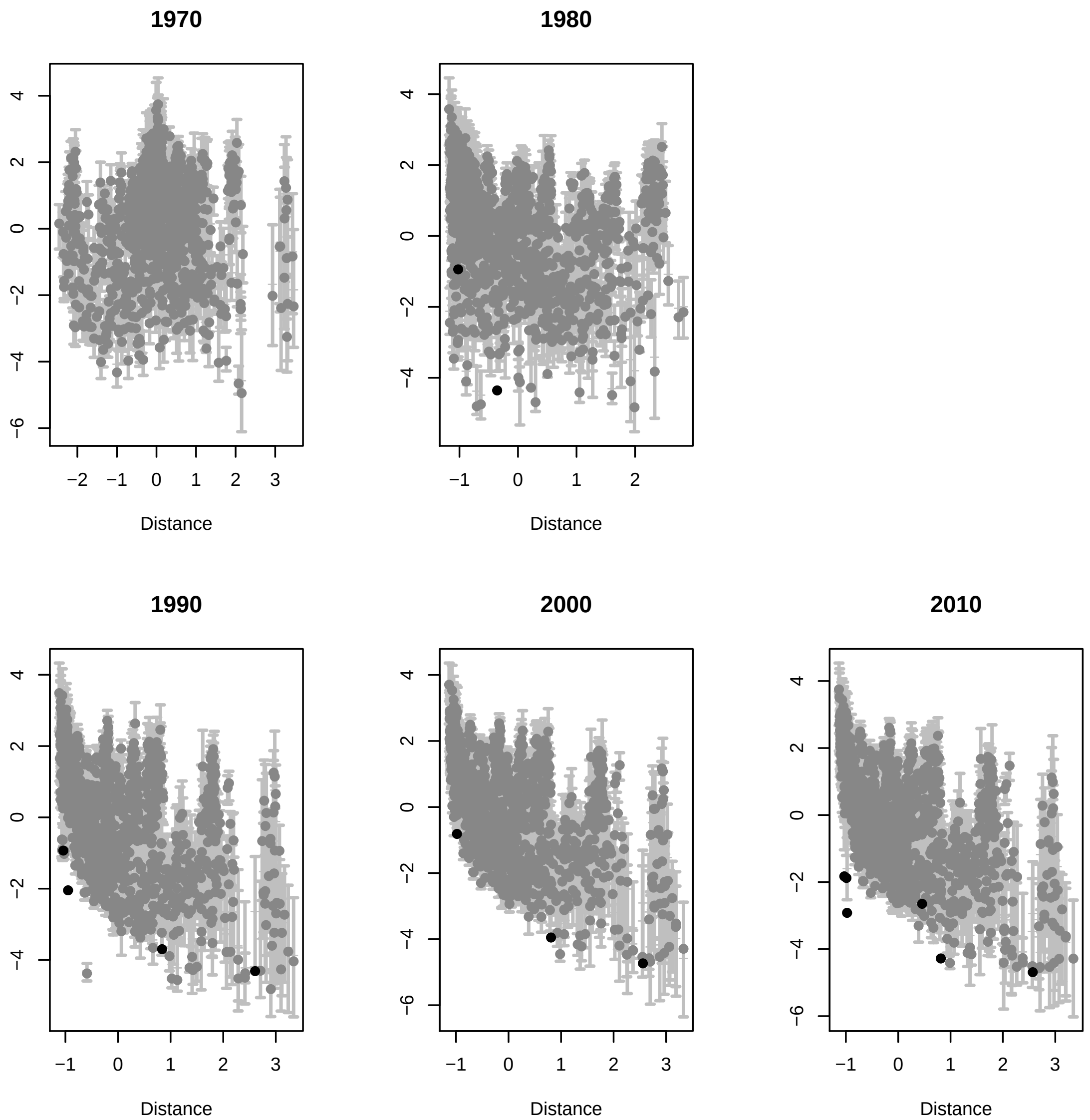


Figure 3: Bayesian 95% credible intervals of the census-tract population log-densities for *Model 7* (light gray vertical bars). Intervals are ranked according to their distance from Boston. Real census-tract population log-densities are denoted by gray circles and Bayesian census-tract outliers by solid black circles.

continue to be weakly attracted by it and experience a lower degree of local spatial dependence.

8 Conclusion

In this paper, we propose a new strategy to model spatial dependence in shaping the population density distribution in Massachusetts over the period 1970-2010. Data suffer from heterogeneity and

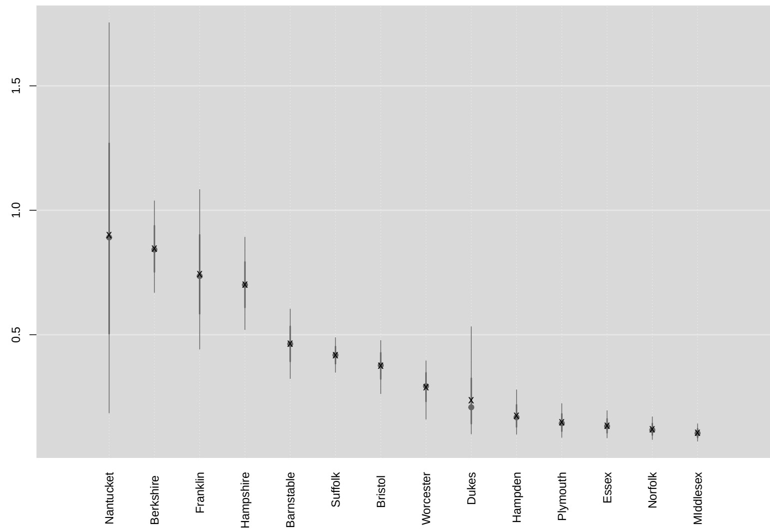


Figure 4: Bayesian 95% credible intervals of the standard deviations $\sigma_1, \ldots, \sigma_{14}$ of the population log-density in each county under *Model 7*. Posterior means are labeled with black crosses and posterior medians with solid gray circles. Counties are sorted horizontally in decreasing order of the posterior medians of $\sigma_1, \ldots, \sigma_{14}$.

heteroscedasticity problems, which we propose to control for with both county and census random effects, the latter being effective in representing local spatial dependence. The principal novelty of our empirical strategy is controlling for local heterogeneity by introducing CAR census effects, along with county random frailties that control for global heterogeneity. The selected model specification confirms that the distance from Boston is the dominant covariate in defining the population density distribution across time. Furthermore, we detect an increasing discrimination effect with an important positive spatial association in modeling population density distribution across time: the ethnic and income characteristics of a neighborhood as well as of the surrounding areas clearly factor into residents' location decisions. However, all of these effects have an important local imprint because CAR spatial dependence among census random effects is a crucial ingredient in guaranteeing the goodness-of-fit of the model. One interesting finding of our contribution is the identification of the importance of spatial dependence at a local level, which has become crucial during the most recent decades. This result witnesses the change in land organization that is taking place in the modern integrated (or globalized) society. Overall, improvements in accessibility, driven by better connections arising from enhanced communication networks, generally seem to rationalize the way activities or agents are distributed across space, entailing spillovers effects involving neighboring territories. Our idea to develop a CAR framework to introduce local spatial dependence in the analysis stems from the advances in other

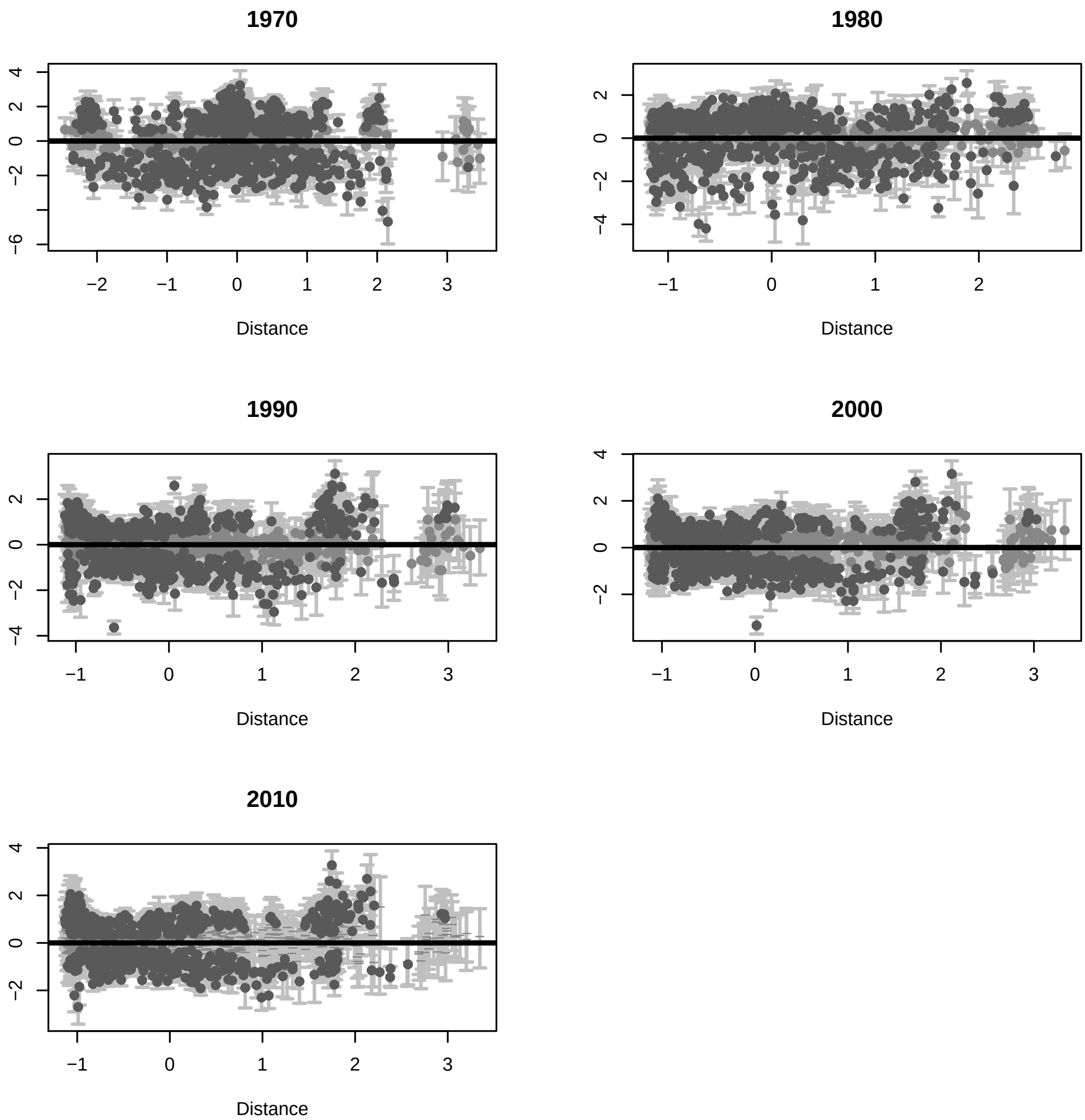


Figure 5: Bayesian 95% credible intervals of the time-independent CAR census random frailties in *Model 7* (light gray vertical bars). Solid gray circles denote the posterior medians of non-null ϕ; solid dark black circles denote the estimated posterior medians of null ϕ.

disciplines. Van Hoef et al. (2018b) discuss that CAR processes are suitable for modeling cases of connectivity among agents in which the network component is an important factor in understanding socioeconomic phenomena. Further advances in this direction would be testing the adoption of our framework of analysis with micro-spatial data in which one is able to control -in a more precise way- for personal relationships in the same spatial units and, above all, for peer effects as incentives that are likely to enforce local spatial dependence. Along the same lines, another interesting extension would be to assess the strength of local spatial dependence for a more micro-spatial setting, for instance, at the census block level. Small spatial units are definitively more suitable for controlling the intensity of individuals' relationships and, hence, for assessing the effectiveness (and magnitude) of local spillovers effects.

From a technical viewpoint, however, our CAR approach suffers from an important limitation: it is unable to control for the temporal evolution of the spatial dependence among census tracts. Because the number of census tracts changes over time, tracking them precisely is not always possible. In that respect, a promising natural extension of our framework is to make it dynamic. The potential availability of panel data is also expected to favor the implementation of the most suitable functional form to jointly model spatial and temporal dependence to enrich our setting and, more precisely, track the evolution of the importance of local spatial dependence. Furthermore, this type of setting would allow for developing a setting of analysis embedding dynamics dimensions and, eventually, being able to extend our framework in the spirit of Costa et al. (2017) so as to explore the potential association between local spatial dependence and population growth.

References

Anselin, L. (2003). Spatial Externalities, Spatial Multipliers and Spatial Econometrics. *International Regional Science Review*, *26*(2), 153-166.

Arab, A., Hooten, M.B., & Wikle, C. (2008). Hierarchical Spatial Models. *Encyclopedia of GIS*, *14*(1), 425-431.

Banerjee, S., Carlin, B.P., & Gelfand, A.E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: CRC Press.

Boarnet, M.G., Chalermpong, S., & Geho, E. (2005). Specification issues in models of population and employment growth. *Papers in Regional Science*, *84*, 21-46.

Browne W.J., Steele F., Golalizadeh M., & Green M.J. (2009). The use of simple reparameterizations to improve the efficiency of Markov chain Monte Carlo estimation for multilevel models with applications to discrete time survival models. *Journal of Royal Statistical Society, Series A*, *172*, 579-598.

Carruthers, J.L., & Mulligan, G. (2008). A locational analysis of growth and change in American Metropolitan Areas. *Papers in Regional Science*, *87*(2), 155-171.

Chi, G., Zhou, X., & Voss, P.R. (2011). Small Area Population Forestacting in an Urban Setting. A Spatial Regression Approach. *Journal of Population Research*, *28* (2), 185-201.

Combes, P.P., & Gobillon, L. (2015). The Empirics of Agglomeration Economies. In G. Duranton, J.V. Henderson & W. Strange (Eds.), *Handbook of Regional and Urban Economics vol. 5* (pp. 247-348). Amsterdam: Elsevier Publisher.

Costa da Silva, D.F., Elhorst, J.P., & da Mota Silveira Neto, R. (2017). Urban and rural population growth in a spatial panel of municipalities. *Regional Studies*, *51*(6), 894-908.

De Olivera, V. (2012). Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics*, *64*, 107-133.

Elhorst, J.P. (2014). *Spatial Econometrics. From cross sectional data to spatial panel.* Heidelberg: Springer.

Elhorst, J.P., Zandberg E., & De Haan, J. (2013). The impact of interaction effects among neighbouring countries on financial liberalization and reform: A dynamic spatial panel data approach. *Spatial Economic Analysis*, *8*(3), 293-313.

Epifani, I., & Nicolini, R. (2013). On the density distribution across space: a probabilistic approach. *Journal of Regional Science*, *53*(3), 481-510.

Epifani, I., & Nicolini, R. (2017). Modelling Population Density Over Time: How Spatial Distance Matters. *Regional Studies*, *51*(4), 602-615.

Epifani, I., & Nicolini, R. (2018). The Importance of Historical Linkages in Shaping Population Density across Space, in Abbruzzo A., Piacentino, D., Chiodi, M. & Brentari, E. (Eds.), *Book of short Papers SIS 2018*, (1196-1201), Pearson.

Gelman, A., Meng, X.J., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, *6*(4), 733-807.

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

Glaeser, E., Scheinkman, J.A. & Shleifer, A. (1995). Economic growth in a cross-section of cities. *Journal of Monetary Economics*, *36*, 117-143.

Glaeser, E., & Ward, B.A. (2009). The causes and consequences of land use regulation: evidence from the Greater Boston. *Journal of Urban Economics*, *65*(3), 265-278.

Halleck Vega, S.H., & Elhorst, J.P. (2015). The SLX model. *Journal of Regional Science*, *55*, 339-363.

Henderson, J.V. (1988). *Urban development, theory, fact, and illusion.* Oxford: Oxford University Press.

Kelsall, J.E., & Wakefield, J.C. (1999). Discussion of "Bayesian models for spatially correlated disease and exposure data" by N. Best, L. Waller, A. Thomas, E. Conlon, & R. Arnold. In J.M. Bernardo, J.O. Berger, A.P. Dawid & A.F.M Smith (Eds.), *Sixth Valencia International Meeting on Bayesian Statistics* (pp. 131-156). London: Oxford University Press.

Jin, X., Carlin, B.P., & Banerjee, S. (2005). Generalized Hierarchical Multivariate CAR Models for Areal Data. *Biometrics*, *61*, 950-961.

LeSage, J.P. (2014). Spatial Econometric panel data model specification. A Bayesian approach. *Spatial Statistics*, *9*, 122-145.

LeSage, J.P., & Pace, K.R. (2009). *Introduction to Spatial Econometrics.* CRC Press, Boca Raton, FL.

LeSage, J.P., & Pace, K.R. (2014). The Biggest Myth in Spatial Econometrics. *Econometrics*, *2*, 217-249.

Majewska, J. (2017). GPS-based measurement of geographic spillovers in tourism- example of Polish districts. *Tourism Geographies*, *19*(4), 612-643.

Maxwell, B.J. (2016). *Exact sparse CAR models in Stan.* `https://mc-stan.org/users/documentation/case-studies/mbjoseph-CARStan.html`.

Quigley, J.M. (1985). Consumer Choice of Dwelling, Neighborhood and Public Services. *Regional Science and Urban Economics*, *15*, 41-63.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, `https://www.R-project.org/`.

Rosenthal, S.S., & Ross, S.L. (2015). Change and Persistence in the Economic Status of Neighborhoods and Cities. In G. Duranton, J.V. Henderson & W. Strange (Eds.), *Handbook of Regional and Urban Economics vol. 5* (1047-1120). Amsterdam: Elsevier Publisher.

Stan Development Team (2017). *RStan: the R interface to Stan. R package version 2.16.2* http://mc-stan.org.

Thomas, A., Best, N.G., Lunn, D., Arnold, R.A., & Spiegelhalter, D.J. (2014). *GeoBUGS User Manual, Version 3.2.3.* Cambridge: MRC Biostatistics Unit.

Topa, G., & Zenou, Y. (2015). Neighbourhood Effects versus Network Effects. In G. Duranton, J.V. Henderson & W. Strange (Eds.), *Handbook of Regional and Urban Economics vol. 5* (561-624). Amsterdam: Elsevier Publisher.

Ver Hoef, J.M., Hanks, E.M., & Hooten, M.B. (2018a). On the relationship between Conditional (CAR) and simultaneous (SAR) autoregressive models. *Spatial Statistics*, *25*, 68-85.

Ver Hoef, J.M., Peterson, E.E., Hooten, M.B., Hanks, E.M., & Fortin, M.J. (2018b). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, *88*, 36-59.

Wall, M.M. (2004). A closed look at the spatial structure implied by the CAR and SAR model. *Journal of Statistical Planning and Inference*, *121*, 311-324.

Wang, X., & Kockelman, K. (2009). Bayesian Inference for ordered response data with a dynamic spatial-ordered probit model. *Journal of Regional Science*, *49*(5), 877-913.

West, M., & Harrison, P.J. (1997). *Bayesian Forecasting and Dynamic Models, 2nd edition.* New York: Springer.

Title Population distribution over time: modeling local spatial dependence with a CAR process

Running Title Population distribution and spatial dependence

Authors

Ilenia Epifani Politecnico di Milano, Dip. di Matematica, P.zza L. da Vinci, 32, I-20133 Milano (Italy), E-mail: ilenia.epifani@polimi.it

Chiara Ghiringhelli Università della Svizzera Italiana, Dip. di Econometria, Via G. Buffi, 13, Lugano, (Switzerland), E-mail: chiara.ghiringhelli@usi.ch.

Rosella Nicolini (Corresponding author) Departament d'Economia Aplicada, Universitat Autònoma de Barcelona, Edifici B – Campus UAB, 08193 Bellaterra (Spain), E-mail: rosella.Nicolini@uab.cat.

Acknowledgements We are grateful to the editor P. Elhorst and two anonymous reviewers for insightful suggestions as well as the participants to NARSC 2016 (Portland) and XII SEA 2018 (Wien) for interesting comments. We also acknowledge the information provided by officers at CIS Mass-State, NAco and Massachusetts Archives. All errors are our own responsibility. This work was supported by the Ministerio de Economia, Industria y Competitividad under Grant ECO2014-52506-R, Generalitat de Catalunya 2017SGR207 and XREPP.

A Appendix: Descriptive statistic on density population in Massachusetts

Table 1 jointly with Figures 1 and 2 track changes of the population distribution in Massachusetts. Table 1 shows that the number of census tracts increased progressively from 1970 to 2010, whereas

	1970	1980	1990	2000	2010
Barnstable	0	0	50	50	56
Hampshire	27	25	30	31	35
Berkshire	15	32	34	41	39
Middlesex	249	271	277	297	317
Bristol	102	105	106	116	125
Nantucket	0	0	4	5	5
Dukes	0	0	4	4	4
Norfolk	101	103	117	121	130
Essex	112	136	146	156	162
Plymouth	46	84	90	90	99
Franklin	0	0	15	16	18
Suffolk	168	177	183	175	193
Hampden	71	83	87	92	103
Worcester	158	157	159	163	171
Massachussets	1049	1173	1302	1357	1457

Table 1: Number of census tracts for each county per decade.

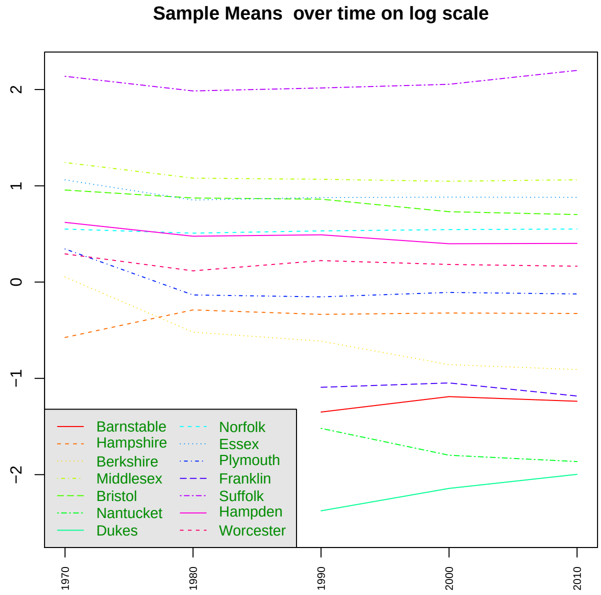
Figure 1 displays that –after 1970– the county population densities (on average) are almost constant over time whereas their standard deviations proportionally shrink, except for the most remote counties.

Furthermore, after 1970, heterogeneity in terms of population size across counties appears jointly with county variance heteroscedasticity. Figure 2 evidences that the greatest concentration of population is in the county of Suffolk which includes the city of Boston.

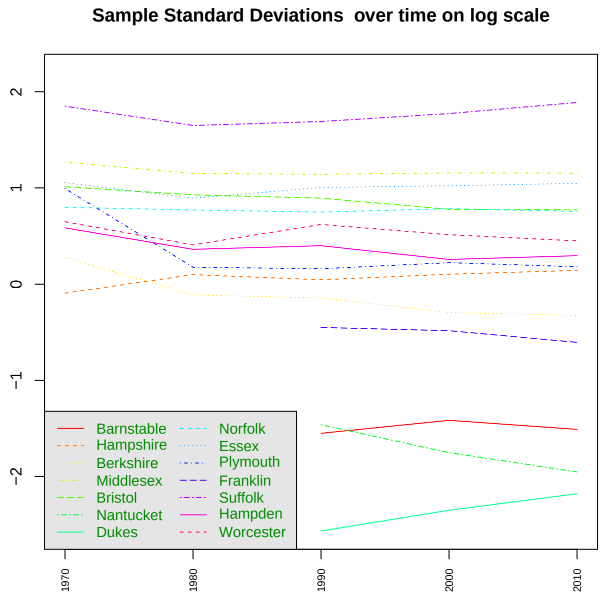
In Table 2, we carry out Kendall's τ index per year as a non-parametric measurement of the association between population density and each covariate. Table 2 depicts that, excluding 1970, the

	1970	1980	1990	2000	2010
Distance from Boston	0.046	-0.275	-0.412	-0.422	-0.441
Ethnic Mix	-0.285	-0.368	-0.468	-0.509	-0.538
Per capita Income		-0.231	-0.226	-0.227	-0.225
Income Gini Index	-0.108	0.412	0.357	0.355	0.364
Education	0.033	-0.214	-0.231	-0.248	-0.239
Amenities	0.189	0.192	0.139	0.145	0.149
Surface	-0.242	-0.255	-0.190	-0.192	-0.195

Table 2: Kendall's τ between density population and each predictor per decade.



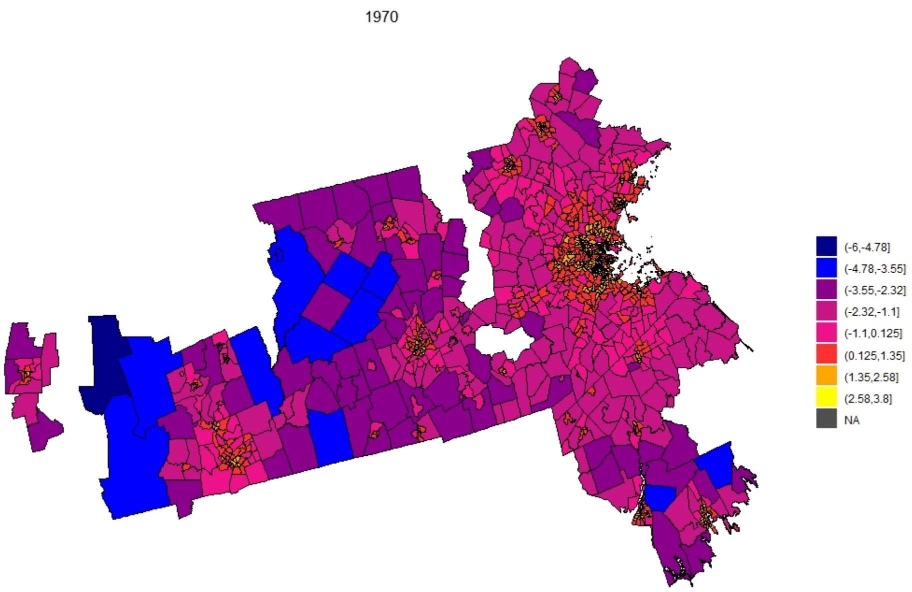
(a)



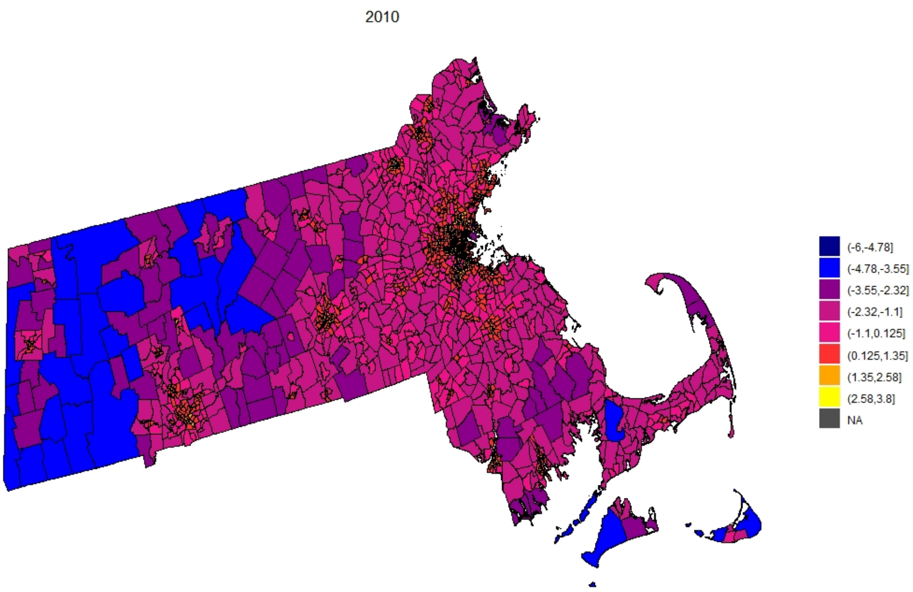
(b)

Figure 1: Sample means in Figure 1(a) and sample standard deviations in Figure 1(b) of the census-tract population densities per square mile of every county in Massachusetts, on a logarithmic scale. Source: US Bureau of the Census; Calculus: authors.

distance from Boston is expected to be relevant in the description of the evolution of the population density distribution, given the strong, negative dependence between *Distance* and density population.



(a)



(b)

Figure 2: Maps of the distribution of the census-tract population densities in Massachusettes for 1970 (in Figure 2(a)) and for 2010 (in Figure 2(b)), on a logarithmic scale. The counties of Barnstable, Nantucket, Duke and Franklin do not appear in 1970 map (white spaces indicate no census tracts in 1970).

As for the other covariates, a negative dependence exists between population density and ethnic *Mix*, and a positive dependence exists between population density and the *Gini* index. Accordingly, we expect that people prefer to move to places (e.g, to live in individual dwellings) where the ethnic composition is distinguished by a high percentage of white people.

B Appendix: Technical details concerning econometric implementation

The Markov Chain Monte Carlo simulation was carried out on a MacBook Pro with CPU at 2.70 GHz with i5 Intel Core, 4 threads, and 16 GB of RAM. All statistical computations and graphics were performed with the R package (R CORE TEAM, 2017). All the models were coded in Stan (Stan Development Team, 2017), which is designed to work closely with the R package. Once the prior distributions and the likelihood have been specified, the code is automatically translated in C++, and a Hamiltonian Monte Carlo routine generates the chains. For each model, we ran a simulation of one chain with 65,000 iterations, 15,000 iterations of burn-in, and a thinning of 10. Therefore, our final sample is made up of 5,000 simulated values. The execution of the simulation took a minimum of 3 hours for *Model 1* and a maximum of 80 hours for *Model 6*.

Following Browne (2009), to improve the performance of the MCMC estimation algorithms in the models including random county effects, we employed a hierarchical centering reparameterization of the county random effects γ_{gt} on the spatial surface S_g and the water areas Z_g. In this respect, S_g, Z_gs are a sort of fixed effects for all census tracts belonging to the same county. The hierarchical centering reparameterization both speeds up the MCMC algorithm and introduces some good mixing benefits in the MCMC generations (i.e., the trace plots are dense, and the autocorrelation plots have fast drop-offs).

As far as the models including CAR census random effects are concerned, we coded exact sparse CAR census random effects as documented in Maxwell (2016) in the following way:

• we implemented a sparse representation of the matrices $\mathbf{W}_t, D_t$ in $(\phi_{1t}, \dots \phi_{n_t t})|\tau_t, \alpha_t \sim$ MN$(\mathbf{0}, \tau_t[D_t - \alpha_t \mathbf{W}_t]^{-1})$ (which scales well for large spatial data sets and expedites computation);

• we manually specified the logarithm of the distribution of the vector $(\phi_{1t}, \dots \phi_{n_t t})$, given τ_t, α_t, directly via the log-probability accumulator (Maxwell, 2016 notes that this trick increases computational efficiency);

• we efficiently computed the determinants of $D_t - \alpha_t \mathbf{W}_t$ using the result

$$det(D_t - \alpha_t \mathbf{W}_t) \propto \prod_{i=1}^{n} (1 - \alpha_t \lambda_{it})$$

where $\lambda_{1t}, \ldots, \lambda_{n_t;t}$ are the eigenvalues of $D_t^{-1/2}\mathbf{W}_t D_t^{-1/2}$, for $t = 1, \ldots, 5$, which can be computed only once and passed in as data. This procedure avoids the computation of $det(D_t - \alpha_t \mathbf{W}_t)$ at each iteration, which is a very computationally demanding operation.

As far as the hyperparameters α_t, τ_t of the CAR census effects $(\phi_{1t}, \ldots \phi_{n_t t})$ are concerned we assumed prior independence such that

$$\text{logit}(\alpha_t) \overset{iid}{\sim} \text{N}(0, 0.02), \quad \tau_t \overset{iid}{\sim} \text{InverseGamma}(0.5, 0.0005) \qquad \forall t = 1, \ldots, T$$

The choice of a logit-normal prior density for the spatial autocorrelation α_t's has been borrowed from `R-INLA` software. This software implements by default that kind of prior for the spatial autocorrelation parameter. First it guarantees that $(\phi_{1t}, \ldots \phi_{n_t t})$ has a multivariate normal density probability at each decade t, as each $\alpha_t \in (0, 1)$ with probability 1. Second, that prior centers α_t at 0.5 (with a prior variance of around 0.0012) without having have heavy tails. In this way, we are assuming a priori the presence of spatial correlation in the model, and this is consistent with the high values of the empirical Moran's I-index of the census tract population log-density per year shown in Table 1. In addition, from a computational viewpoint, a logit-normal$(0, 0.02)$ (rather than the more heavy-tailed Uniform$(0, 1)$) prevents the simulation of α_t from being stuck near the boundary one, without limiting its point estimate to approach it, if so demanded by data. In general, heavy-tailed distributions make Stan perform poorly since sampling from them is difficult for the Hamiltonian Monte Carlo algorithm (Gelman, 2019 in `https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations`).

For variances τ_t's, the prior InverseGamma$(0.5, 0.0005)$ distribution guarantees that all reasonable levels of variability are considered while avoiding the exclusion of small values. As Kelsall and Wakefield (1999) have posited, in case of CAR random effects, the most common prior InverseGamma$(0.001, 0.001)$ tends to place most of the prior mass away from zero (on the scale of the standard deviation). In situations in which the true spatial dependence between areas is negligible (i.e., with a standard deviation close to zero), this dynamic may generate an artefactual spatial structure in the posterior. For that reason, following Kelsall and Wakefield (1999), we set the InverseGamma$(0.5, 0.0005)$ prior distribution for τ_t. According to that assumption, the prior belief on the standard deviation of the CAR census random effects centers around 0.5 with a 1% prior probability of being less than 0.01 or larger than 2.5 (refer to the GeoBUGS User Manual by Thomas et al. 2014).

The convergence diagnostics (Geweke test, effective sample size, traceplot, autocorrelation function) were computed for all parameters, indicating that convergence was achieved for almost all of them.

C Appendix: Bayesian methods for goodness-of-fit and model selection

We assessed the goodness-of-fit of our model specifications by computing the Bayesian posterior p-values of the alternative econometric specifications on the basis of a discrepancy measure between the data of population density $\boldsymbol{y}$ and each *Model M*, for $M = 1, \ldots, 7$. We adopted the χ^2 discrepancy measure suggested in Gelman et al. (1996):

$$\chi^2(\boldsymbol{Y}^{\text{rep}}, \boldsymbol{\theta}_M) = \sum_{i,t} \frac{\left(\log(Y_{it}^{\text{rep}}) - \text{E}(\log Y_{it}^{\text{rep}} | \boldsymbol{\theta}_M)\right)^2}{\text{Var}(\log Y_{it}^{\text{rep}} | \boldsymbol{\theta}_M)}$$

where $\boldsymbol{\theta}_M$ summarizes all the unknown parameters in *Model M*, and $\boldsymbol{Y}^{\text{rep}}$ denotes the replicated data from the posterior-predictive distribution $f(\boldsymbol{Y}^{\text{rep}} | \boldsymbol{Data}, M)$ under *Model M*. Hence, the Bayesian posterior p-value p_M of *Model M* is defined as the tail area posterior-predictive probability that the discrepancy $\chi^2(\boldsymbol{Y}^{\text{rep}}, \boldsymbol{\theta}_M)$ exceeds the "realized" value $\chi^2(\boldsymbol{y}, \boldsymbol{\theta}_M)$, given all $\boldsymbol{Data}$, that is,

$$p_M = P\left(\chi^2(\boldsymbol{Y}^{\text{rep}}, \boldsymbol{\theta}_M) > \chi^2(\boldsymbol{y}, \boldsymbol{\theta}_M) | \boldsymbol{Data}\right) \tag{1}$$

Values of p_M around 0.5 indicate that model M fits quite well, while values close to zero or one suggest a limited fitting capacity. Computation of p_M has been performed via Markov Chain Monte Carlo simulation.

A Bayesian comparison of the alternative specifications for *Model 1–7* was performed by computing the Bayesian Deviance Information Criterion (DIC), the Logarithm of the Pseudo-Marginal Likelihood, and the Percentage of the Bayesian Census Outliers. The Bayesian DIC is a generalization of the Akaike's Information Criterion (AIC), and it is given by the deviance (i.e., two times the log-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 to those with a large DIC. For each decade t, the Logarithm of the Pseudo-Marginal Likelihood ($LPML$) is defined as the sum of the logarithms of the *Conditional Predictive Ordinates* (CPO), and each CPO_{it} is given by the value of the posterior-predictive population density evaluated at the actual y_{it}, conditionally to the sample

$\boldsymbol{y}_{-it}$ not containing any data from census tract i at decade t. Thus,

$$LPML_t = \sum_{i=1}^{n_t} \log(CPO_{it}) \,.$$

The larger the value of the CPO's (and hence the larger the value of the $LPML$), the better the fit of the model.

Last, a census tract i at time t is defined to be a *Bayesian Outlier* with level 90% if its real population density y_{it} falls into one of the two 5% tails of the marginal posterior-predictive density.

D Appendix: Map of Massachusetts color-coded according to the posterior CAR census frailties

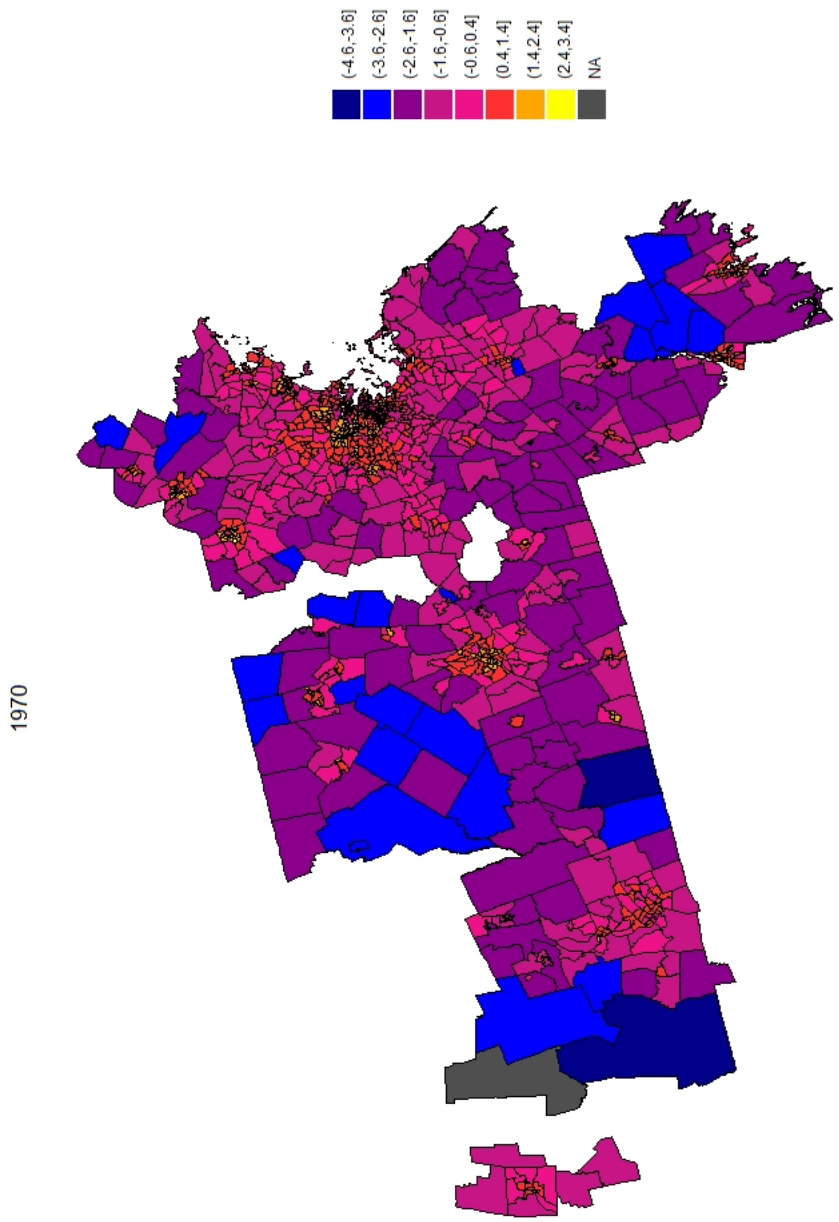


Figure 3: Counties in Massachusetts in 1970 color-coded according to the estimated posterior median values of the CAR census frailties under *Model 7*. The counties of Barnstable, Nantucket, Duke and Franklin do not appear in this map (white spaces indicate no census tracts in 1970, 1980).

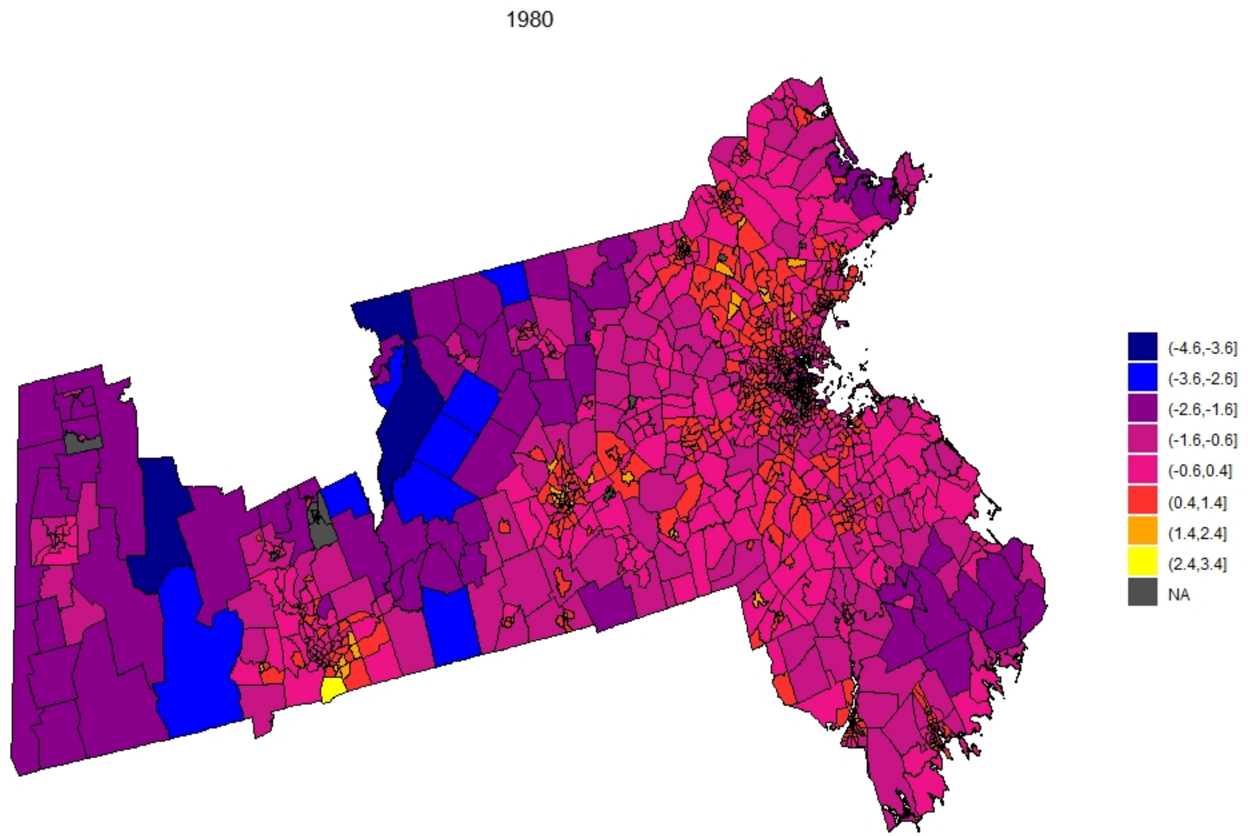


Figure 4: Counties in Massachusetts in 1980 color-coded according to the estimated posterior median values of the CAR census frailties under *Model 7*. The counties of Barnstable, Nantucket, Duke and Franklin do not appear in this map (white spaces indicate no census tracts in 1970, 1980).

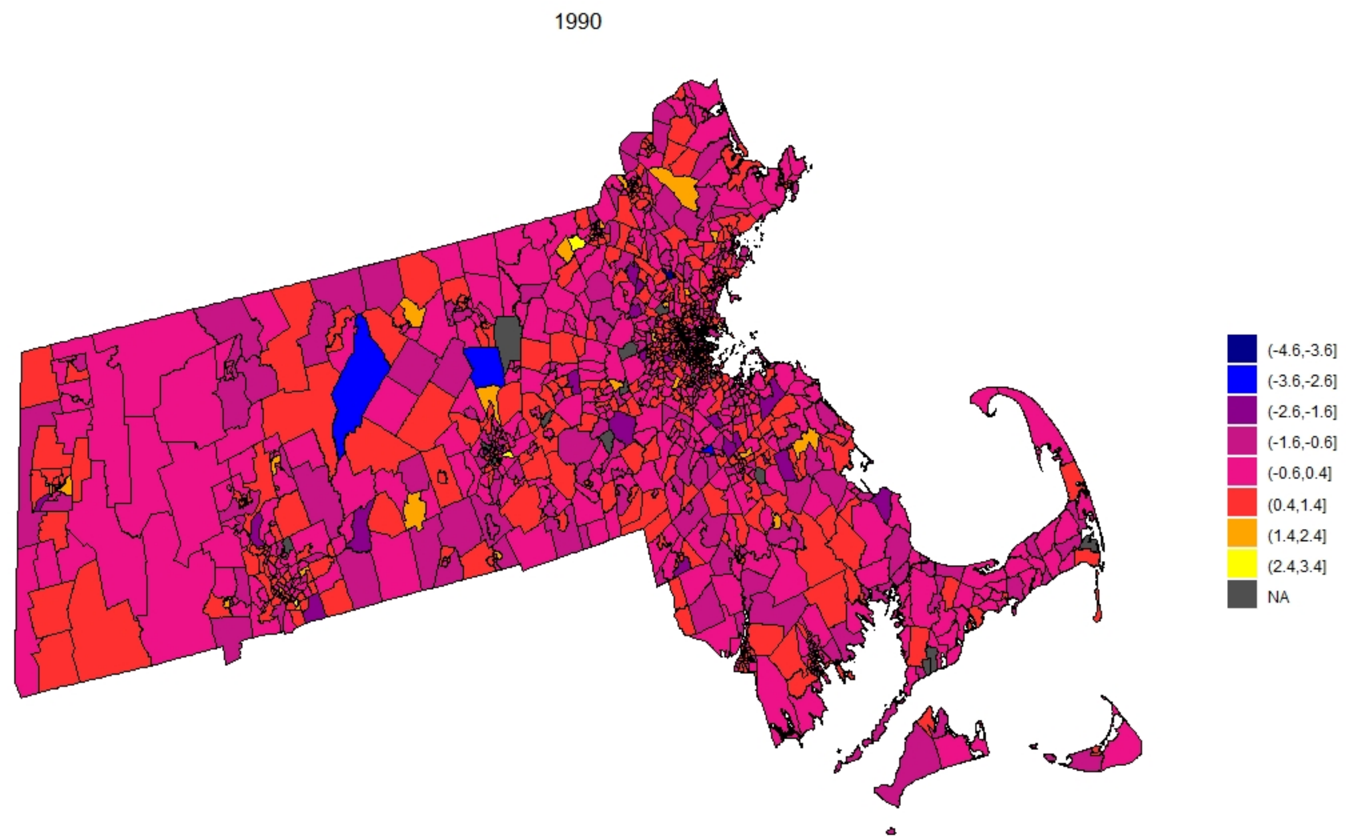


Figure 5: Counties in Massachusetts in 1990 color-coded according to the estimated posterior median values of the CAR census frailties under *Model 7*.

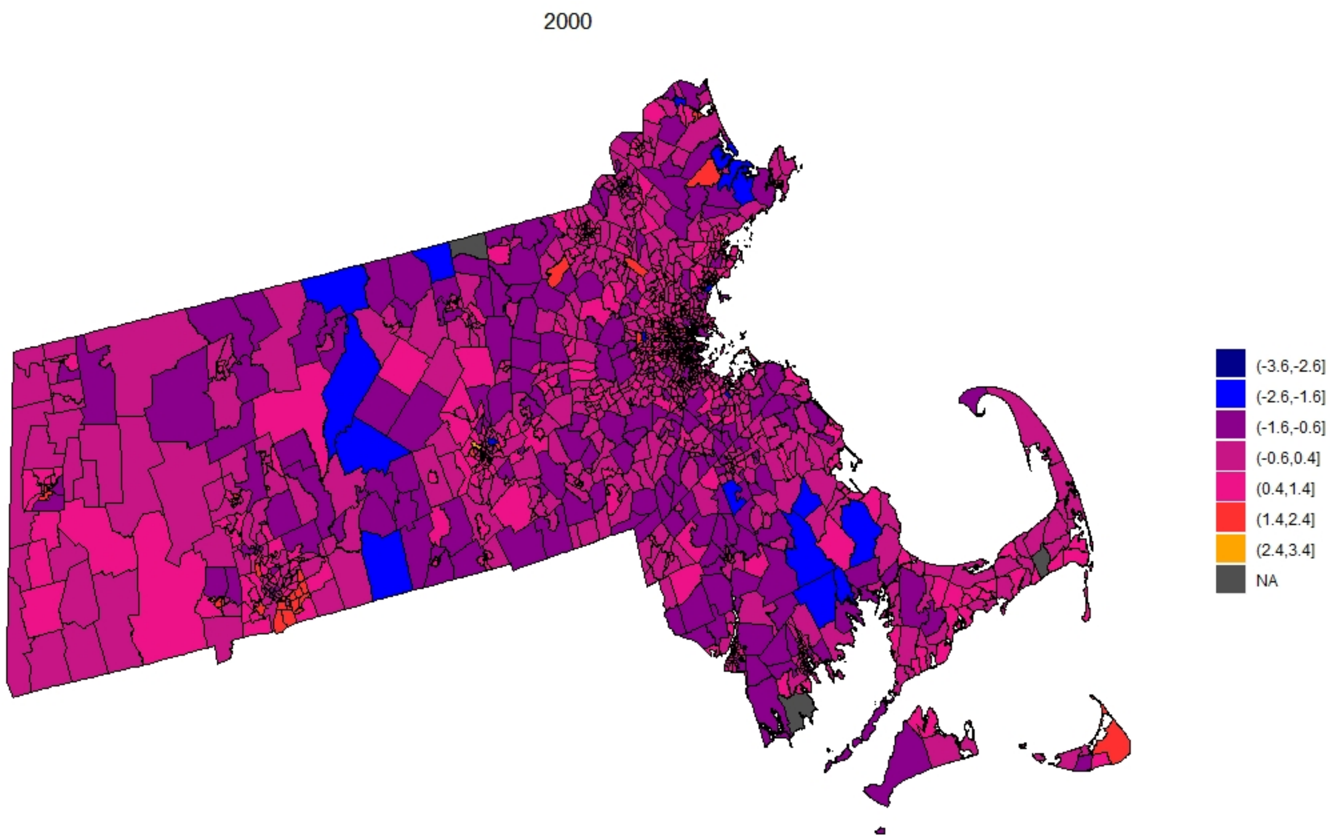


Figure 6: Counties in Massachusetts in 2000 color-coded according to the estimated posterior median values of the CAR census frailties under *Model 7*.

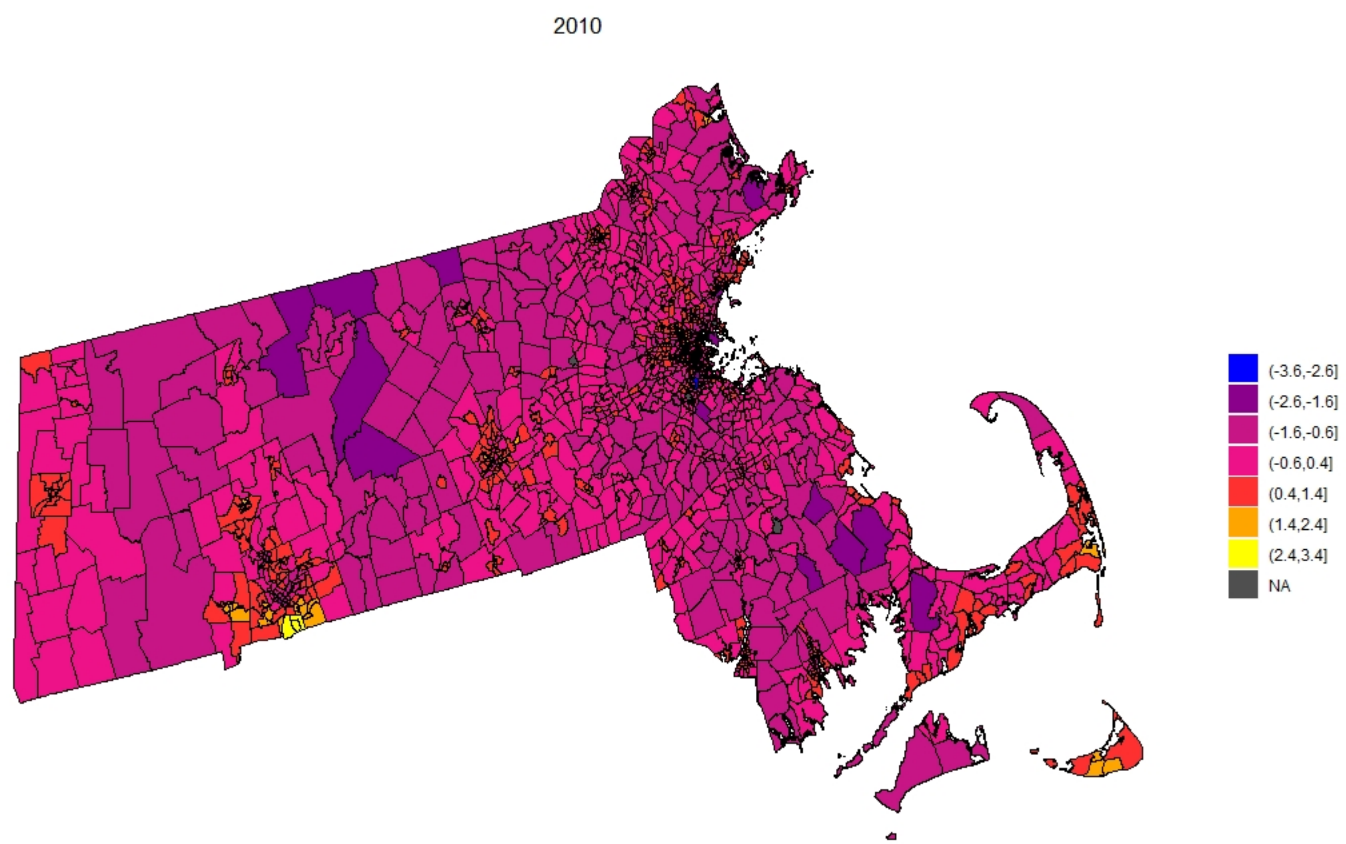


Figure 7: Counties in Massachusetts in 2010 color-coded according to the estimated posterior median values of the CAR census frailties under *Model 7*.