Accepted Manuscript

Proximity effects in obesity rates in the US: A Spatial Markov Chains approach

Massimiliano Agovino, Alessandro Crociata, Pier Luigi Sacco

PII: S0277-9536(18)30647-6

DOI: https://doi.org/10.1016/j.socscimed.2018.11.013

Reference: SSM 12033

To appear in: *Social Science & Medicine*

Received Date: 5 May 2018

Revised Date: 2 September 2018

Accepted Date: 6 November 2018

Please cite this article as: Agovino, M., Crociata, A., Sacco, P.L., Proximity effects in obesity rates in the US: A Spatial Markov Chains approach, *Social Science & Medicine* (2018), doi: https://doi.org/10.1016/j.socscimed.2018.11.013.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Proximity effects in obesity rates in the US:

A Spatial Markov Chains approach

Massimiliano Agovino[d], Alessandro Crociata[e] and Pier Luigi Sacco[a,b,c,f*]

[a] IULM University, Via Carlo Bo, 1, 20143 Milan, Italy. pierluigi.sacco@iulm.it

[b] Harvard University, Boylston Hall, Harvard Yard, Cambridge, MA 20138 USA pierluigi_sacco@fas.harvard.edu

[c] metaLAB (at) Harvard, Cambridge, MA USA

[d] Parthenope University, Naples, Italy

[e] Gran Sasso Science Institute - GSSI, L'Aquila, Italy

[f] FBK-IRVAPP, Trento, Italy

* Corresponding author

The data used in this paper are in the public domain and freely accessible at http://www.americashealthrankings.org/explore/

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Proximity effects in obesity rates in the US:

A Spatial Markov Chains approach

Abstract

In this paper, we investigate, by means of a Spatial Markov Chains approach, the existence of proximity effects at State level for US data on obesity rates in the period 1990-2011. We find that proximity effects do play an important role in the spatial diffusion of obesity (the obesity 'epidemics'), and that the actual health geography of nearby States in terms of high vs. low obesity rates makes an important difference as to the future evolution of the State's own obesity rate over time. This means, in particular, that clusters of States characterized by uniformly high levels of obesity rates, as it happens for instance in the US Southern macro-region, may suffer from a perverse 'geographical lock-in' effect that calls for coordinated action across States to implement effective countervailing policies.

Keywords

Obesity rates; obesity epidemics; proximity effects; Spatial Markov Chain; Ergodic distribution.

ACCEPTED MANUSCRIPT

1. Introduction

In view of the rapid propagation of obesity throughout human populations at a global scale, there is an increasing tendency to speak not only of 'obesity epidemics' (Contaldo and Pasanisi, 2004, 2005), but even of a global pandemic (Swinburn et al., 2011). Moreover, morbid obesity has been found to be increasing at a fast rather than moderate pace (Sturm, 2007). The projected global trends are particularly alarming: A recent, very large scale study with almost 20 million participants predicts that, according to the current trend, global obesity prevalence will amount to 18% in men and to more than 21% in women by 2025 (NCD-RisC, 2016), with dire prospects in terms of health consequences and welfare cost burden (Wang et al., 2011). The emphasis on the 'epidemic' character of obesity seems to reflect a widespread idea that the role of a variety of social and cultural factors in determining obesity-related habits and attitudes may be bigger than previously understood, and is fundamentally transmitted and stabilized through the structure of social ties (Christakis and Fowler, 2007), and more generally through various forms of social interaction (Santonja et al., 2010; Ejima et al., 2013).

The debate on the causes of this phenomenon is still very lively. A recent review by Heymsfield and Wadden (2017) emphasizes the importance of the combined effect of high-calorie palatable foods supplied in large portions and the increased incidence of both occupational and leisurely sedentary activities. The role of the availability of cheap, inviting high-caloric food coupled with socially influenced failures of individual self-control in food choices (Elfhag and Morey, 2008; Sacco, 2017) and physical activity habits (Sniehotta et al., 2005) seems therefore particularly relevant. This line of explanation is compatible in principle with those highlighting the combined role of evolutionary mismatch and socioeconomic inequality (Hruschka, 2012), and it seems to command more scholarly consensus than alternative explanations based upon food insecurity (Nettle et al., 2017). Although there is not a general enough consensus on basic theoretical explanations yet (Mullan et al., 2017), there may be some conceptual and practical advantage in studying the combined role of social and geographical factors in the onset and propagation of obesity in terms of a (pseudo-) epidemic process.

Among the socio-geographical factors that are more invoked in the literature there is the phenomenon of the so called 'food deserts', i.e. the fact that socially deprived neighborhoods may be lacking delivery points of healthy affordable food, thus favoring the diffusion of bad eating habits in the poorer segments of the population (Jetter and Cassady, 2006; Morland and Evenson, 2009). However, there are still many issues with the definition and operationalization of the 'food desert' notion (Walker et al., 2010), and the differences in terms of effective nudging of healthy food choices at delivery points in high-income vs. low-income areas seems less linked to physical reachability and availability of healthy choices than to in-store display and marketing choices for items on sale (Ghosh-Dastidar et al., 2014), and therefore, as already mentioned, to issues of self-control (Lawrence et al., 2012). Rather than 'food deserts', it seems that 'food swamps' (i.e. delivery points with mostly unhealthy food options) should therefore be seen as the major threat to healthy eating habits (Cooksey-Stowers et al., 2017). The emphasis seems then to shift away from locational scarcity factors in favor of economic factors such as affordability (Inglis et al., 2009; Lin et al., 2014), and behavioral factors such as promotion and palatability of healthy choices vis-à-vis unhealthy ones (Hawkes et al., 2015). There has been a stream of literature that has tried to conceptualize obesity as the result of rational economic choice for certain combinations of resources and incentives (Chou et al., 2004). However, keeping into account the interaction between economic incentives and affective factors (Ruhm, 2012) or social pressure (Dragone and Savorelli, 2012), it turns out that undesired overeating, and therefore obesity, may be far from unlikely outcomes also for rational decision makers.

ACCEPTED MANUSCRIPT

Even if some elements of rational decision making may be at play in food choices, and specifically in the genesis of obesity, the role of social pressure and incentives seems therefore at least as strong. Research on adolescent behavior related to the formation of food habits clearly confirms the role of homophily and peer pressure in determining similar adiposity levels across friendship networks (Renna et al., 2008; Trogdon et al., 2008; Valente et al., 2009; Halliday and Kwak, 2009; de la Haye et al, 2010). Similar effects are found for obesity-relevant habits such as attitude to physical activity (de la Haye et al., 2011; Fitzgerald et al., 2012).

In this paper, we attempt at reconciling the socio-geographically and socio-behaviorally inspired streams of literature on obesity by taking the epidemics metaphor seriously in explaining obesity diffusion patterns. Social diffusion is the product of the complex interaction of different mechanisms at different spatial scales, from social contagion to homophily (Rohilla Shalizi and Thomas, 2011). Explicit 'contagion' models for the diffusion of obesity have been developed (Cohen-Cole and Fletcher, 2008). At the same time, the spatially characterized differences in socioeconomic status in all kinds of residential environments are clearly a possible factor of diffusion of eating behaviors and habits (Shahar et al., 2005), and have even been tracked in their effects on obesity rates through a randomized social experiment (Ludwig, 2011); see Arcaya et al. (2016) for an up-to-date review of the research on neighborhood effects and health. This basic factor is further compounded with other intervening variables such as age (Baum and Ruhm, 2009), ethnicity (Scharoun-Lee et al., 2009), gender (Dammann and Smith, 2009), or a combination of the above (Zhang and Wang, 2004), as well as with local community variations in collective efficacy (Cohen et al., 2006), social influence (Zhang et al., 2015), or social norms (Shoham et al., 2015). Moreover, socioeconomic differences even influence the impact of obesity on health-related quality of life (Minet Kinge and Morris, 2010).

Trying to model this complex web of social, cultural and economic influences into a comprehensive micro-social model is very difficult. However, as these factors all interact in determining socially mediated patterns of spatial diffusion of attitudes and behaviors, it is possible to try and model their aggregate influence at the macro level in terms of proximity effects. The existence of proximity effects across neighboring States finds support in the US in the light of recent research by Nelson and Rae (2016), that shows how the US may be split into inter-state megaregions on the basis of major commuter patterns, thus defining an emergent geography of social transmission. The structure of such megaregions shows that mobility-driven, systematic interaction between neighboring States is the norm, thus providing, in principle, an empirical basis for a proximity-based model of obesity epidemics at State level. Moreover, many health-related policies and regulations, such as in the case of childhood obesity measures, are defined at State level, making the latter a meaningful territorial level of analysis of the obesity epidemics (Dodson et al., 2009).

To our knowledge, empirically driven research on proximity effects in social medicine is still lacking, and our methodological approach is novel in the field. In this paper, we explicitly model the obesity epidemics phenomenon in terms of a proximity effect dynamics, by means of a Spatial Markov Chains approach on US data at State level, 1990-2011. We find that obesity rates are clearly subject to proximity effects at the State level, thus suggesting that a spatial diffusion dynamics is at work, and that therefore the idea that obesity may be 'socially contagious' is compatible with the available evidence. To our knowledge, this is the first example of data-driven modeling of obesity epidemics in terms of a diffusion model. Our findings could inspire a new generation of obesity prevention strategies that keep into account the spatial patterns of social transmission mechanisms in their geographical variations.

In section 2, we briefly introduce the Spatial Markov Chains approach. In section 3 we present the data and some preliminary findings. In section 4 we illustrate the main results. Section 5 briefly discusses them and concludes. A short technical Appendix closes the paper.

ACCEPTED MANUSCRIPT

2. Method: The Spatial Markov Chains Approach (SMCs)

Spatial econometrics has been a rapidly expanding area of research in the last decade (Griffith and Paelinck, 2007; Anselin, 2010) to study the socio-spatial dynamics of a wide variety of phenomena. Applications in social medicine, however, are still limited to date. In this paper, we make use of spatial econometric techniques to model the diffusion of obesity as a social epidemics phenomenon. In particular, we work with Spatial Markov Chains (SMCs), which allow us to simultaneously analyze the spatial and time dynamics of the process. In this paper, we make use of the classical SMCs methodology as developed in the seminal papers of Rey (2001) and Le Gallo (2004). Such methodology has been used in a variety of different fields, such as the dynamics of regional wealth disparities (Yue et al., 2014), the diffusion of pro-environmental behaviors (Agovino et al., 2016), the evolution of regional competitiveness in manufacturing (Schettini et al., 2011), and so on. On the other hand, to our knowledge there has been so far no research on proximity effects in obesity epidemic phenomena. Research on the spatial determinants of obesity has mainly concentrated on the relationship between location of food delivery points and eating behaviors (Davis and Carpenter, 2009), or between obesity and distance from recreational areas (Wolch et al., 2011).

The main output of a SMCs model is the spatial transition matrix, which evaluates how nearby locations at a given scale influence each other as to the observed levels of the variables under study. In our specific context, the matrix measures the positive or negative influence of neighboring States on the transition of a given US State across different levels of health variables on a given measurement scale (e.g., good, fair, average, inadequate and bad). In particular, the matrix provides, for a given State, the probability to move upwards or downwards in the distribution in the next period (*t+1*), conditional upon the state of its neighbors in the current period (*t*). The transition matrix can therefore trace the history of the distribution of values over time. More specifically, this technique allows us to track whether a State characterized by an unsatisfactory (satisfactory) obesity rate tends to remain in that status if surrounded by other States with similarly unsatisfactory

(satisfactory) obesity rates, and in particular whether States with unsatisfactory obesity rates negatively affect their neighbors, pulling up their obesity rates, or likewise States with satisfactory obesity rates positively influence their neighbors, by pulling down their obesity rates. Proximity effects can be understood as the aggregate result of social transmission processes which have maximum intensity at the local scale and decay with distance (Madan and Pentland, 2009); for an explicit modeling of the micro-social diffusion dynamics of obesity-related attitudes and behaviors see Madan et al. (2010). We can therefore build a dynamic model that analyzes the evolution of these proximity effects over time, and test it on available observed data.

The building of the spatial transition matrix is based upon the traditional Markov transition matrix, that yields the spatial transition probability. In particular, in this approach the traditional transition matrix is modified so that the transition probabilities of a State in the next period *(t+1)* are conditional upon the average level of the obesity rates at the current period (*t*) in its neighboring States. In other words, the SMCs spatial transition matrix expands a *K-by-K* traditional Markov transition matrix into *K* conditional matrices[1], each of which is in turn a *K-by-K* matrix. In our case, *K* is equal to 5, the number of possible classes, that is, for each possible class we have a corresponding conditional matrix (see Table 1 below for an example).

In formal terms, if we consider the *k-th* matrix among the conditional matrices, the $p_{ij}(k)$ element of such a matrix represents the probability that a State located in class *i* in the current period (*t*) ends up in class *j* in the next period (*t+1*), knowing that the average level of the obesity rates in its neighboring States belongs to class *k* in period *t*. The $p_{ij}(k)$ element of a conditional transition matrix is thus defined as follows:

$$\hat{p}_{ij}(k) = \frac{n_{ij}(k)}{n_i(k)}$$

[1] A conditional matrix is a matrix in which the probability of the (health) status of a State at time (t + 1) is conditional upon the (health) status of its neighbors at time t.

where $n_{ij}(k)$ is the number of States located in class i in period t, and in class j in period ($t+1$), conditional upon an average level of obesity rates in neighboring States belonging to class k in period t. $n_i(k)$ is the total number of States belonging to class i, conditional upon an average level of obesity rates in neighboring States belonging to class k at time t, for T=21 annual transitions[2], i.e. $n_i(k) = \sum_j n_{ij}(k)$.

The spatial Markov matrix allows us to appreciate the positive or negative influence of the neighbors on the transition of a State across levels of obesity rates. Indeed, the influence of spatial proximity effects is reflected in the differences between the unconditional[3] transition values and the conditional ones (Le Gallo, 2004). For example, in our case with 5 classes ($K=5$), the first class groups States with the best health status (low obesity rates), the third corresponds to States with intermediate health status, and the fifth to States with the worst health status (high obesity rates)[4]. Consequently, if $p_{AB}>p_{AB|G}$, the transition probability of moving 'upwards' (i.e. increasing its obesity rate) for a State with an intermediate obesity rate without proximity effects, i.e. not taking into account the social transmission effects associated to its neighbors' obesity rates, is larger than the transition probabilities of moving 'upwards' for a State with an intermediate obesity rate conditional upon neighbors with the lowest obesity rates (notice that moving 'upwards' here means lowering the obesity rate, i.e. improving the health status, and accordingly for moving 'downwards'). Likewise, if we consider the probability of moving 'upwards' for States starting from different classes of obesity rates. Conversely, if $p_{GA}<p_{GA|B}$, the transition probability of moving 'upwards' for a State with a low obesity rate conditional upon neighbors with high obesity rates is larger than the transition probabilities of moving 'upwards' for a State with a low obesity rate in the absence of proximity effects. **Table 1** summarizes the reasoning and offers a key to understanding the Spatial Markov Chains approach.

[2] Our period of analysis consists of 22 years, so we have $T = 21$ annual transitions.
[3] For reasons of space, we do not report the values of the unconditional transition matrix. Interested readers can request them to the authors.
[4] We describe the 5 classes more in detail in Section 4.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[Insert Table 1 about here]

If proximity effects do not matter for transition probabilities, then the conditional probabilities should be equal to the unconditional initial transition values (Le Gallo, 2004):

$$p_{ij|1} = p_{ij|2} = \cdots = p_{ij|5}$$

$$\forall i = 1, \dots, 5$$

$$\forall j = 1, \dots, 5$$

The relevance of the socio-spatial dimension, and therefore the importance of considering neighbors in determining transition probabilities, corresponds to the rejection of the null hypothesis of spatial stationarity tests (see Le Gallo, 2004). On the basis of our data (see below), we reject the null hypothesis at 5% and, consequently, the transition probability of a State does depend on the spatial environment, so that proximity effects matter.

3. Data and preliminary results

In this section, we introduce the obesity rates data at US State level (section 3.1), and subsequently we carry out a preliminary analysis (section 3.2). In particular, we check whether proximity effects are relevant as to the influence of the obesity rate in a given State upon the obesity rates in nearby ones, or on the contrary whether the obesity rate evolves independently of those in neighbor States. In other words, we ask whether it is possible to identify spatial dependence patterns so that a State's obesity rate interacts with those of neighboring States through social transmission effects. Our analysis is based upon two statistical tools, the Theil index (TI), and the Moran index (MI), according to the conceptual framework introduced below (Section 3.2).

3.1 Data used and data limitations

In our analysis, we use America's Health Rankings data for the 51 US States, over the period 1990-2011. Presented by the United Health Foundation[5], the America's Health Rankings Annual Report has tracked the health of the nation for 26 years, providing a specific, comprehensive data environment on the evolution of the health status in the USA at State level.

OR is the percentage of adults who are estimated to be obese, defined as having a body mass index (BMI) of 30.0 or higher, according to self-reported height and weight. BMI is equal to weight in pounds divided by height in inches squared and then multiplied by 703. The Center for Disease Control (CDC) has a calculator for BMI. Because of the 2011 change in BRFSS (Behavioral Risk Factor Surveillance System) methodology, obesity prevalence from the 2012 Edition onward cannot be directly compared to estimates from previous years[6].

Figure 1 is built on the basis of the regional divisions used by the United States Census Bureau (i.e., Northeastern, Midwestern, Southern, and Western States). As it was anticipated in the literature review, the evidence shows that the obesity rate follows a growing trend, both in the aggregate and for each region. In particular, we note that the Midwestern and Southern US are the regions with the highest obesity rates, exceeding national ones. On the contrary, Western and Northeastern US are the regions with the lowest obesity rates, much lower than the national ones.

This result is particularly evident if we look at the quartiles maps of the obesity rates (Figure 2). In particular, we find a clear clustering tendency over the years. If in 1990 the obesity rate appears as a spatially disperse social phenomenon, looking like a fragmented patchwork on the national map, in the subsequent five-year intervals we notice that Southern States gradually group together, and emerge as those with the highest obesity rates, whereas, likewise, Western States similarly coalesce as the States with the lowest obesity rates. The time evolution of obesity rates at the national scale

[5]The United Health Foundation was established by UnitedHealth Group in 1999 as a not-for-profit, private foundation dedicated to improving health and health care (http://www.unitedhealthfoundation.org/AboutUs/Default.aspx).
[6] http://www.americashealthrankings.org/WV/Obesity

as it is readable from the sequence of maps in Figure 2 suggests that some social transmission mechanism is at work, causing a spatial diffusion process that exacerbates difference in obesity rates at the regional level.

In our analysis, we use data at State and not county level. The State-level geographical scale might be deemed too big for the study of the social diffusion of obesity (although, as already noted, this is the territorial level at which most policy measures are defined). Clearly, data at county level are more fine-grained in their empirical modeling of the spatial element. Moreover, it is well known that administrative data aggregate individuals on the basis of arbitrary geographical boundaries, that reflect political and historical situations (see Arbia, 2005; Arcaya et al., 2012). The choice of the spatial aggregation unit is therefore essential, as different choices may lead to different results in the estimates (see Rey, 2001). Data at the State level cannot be considered as “independently generated” (Anselin 1988; Anselin and Bera 1998) because of spatial similarities of neighboring States; thus, standard estimation procedures can provide biased estimators of the parameters. Aggregating data at the county level would allow spatial effects, such as spatial spillovers, to be more properly modelled in principle (Arbia et al., 2002; Arbia, 2005; Agovino et al., 2016).

One way to overcome this problem could have been building our analysis upon county-level data, as provided by the Center for Disease Control and Prevention[7]. Unfortunately, however, such data are only available for the period 2004-2013, and this would have cut our time series down by 11 years (State level data are available for the period 1990-2011), seriously limiting the results of the empirical analysis conducted through the Spatial Markov Chains approach, which is substantially affected by the size of the historical data series (see Rey, 2001). In particular, such analysis returns the long-term distribution of the studied phenomenon (ergodic distribution; see Section 4), and a sequence of 10 years is too short to study the long-term distribution of a social phenomenon. Finally, the use of data at State level could weaken the spatial correlation, as larger administrative

[7] https://www.cdc.gov/diabetes/data/countydata/countydataindicators.html (last access July 2018).

units have the drawback of reducing the heterogeneity that is present in finer data (Arbia, 2005). It follows that an analysis of data at the county level would return a relatively stronger spatial correlation and clustering, with stronger spatial and temporal persistence. On the other hand, in the light of the above, if substantial levels of spatial correlation and clustering are already found with State-level data, this qualifies as strong evidence for the existence of spatial effects.

[Insert Figure 1 about here]

[Insert Figure 2 about here]

3.2 Preliminary results

In this section, we conduct a preliminary analysis by means of the joint application of a measure of inequality and of the degree of spatial autocorrelation. We recur to two different indexes, i.e. the Theil and Moran indexes, whose simultaneous use provides complementary analytic insights, which cannot be obtained if they are used separately (see Rey, 2001).

Gezici and Hewings (2007) highlight the relevance of the joint use of inequality and spatial indices, especially in the study of socio-economic phenomena characterized by persistent spatial clustering processes over time. In particular, the Theil index is a measure of total inequality, and, in the context of our study, can be defined as:

$$\mathrm{TI} = \sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{d_i} \log(\mathrm{pd_i}) \qquad (1)$$

where p is the number of States, and $\mathrm{d_i} = \mathrm{OR_i} / \sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{OR_i}$ whereas OR_i is OR at State level[8]. TI takes on values in the interval *[0; log(p)]*, and is, in particular, equal to 0 in the case of a perfectly even spatial distribution, and to the highest value, *log(p)*, when OR is entirely concentrated in a single State.

[8] We consider the logarithm of OR.

ACCEPTED MANUSCRIPT

We also consider the Moran index (MI) as a measure of the degree of spatial autocorrelation, which has the following definition:

$$MI = \frac{z_k^{'} W z_l}{z_k^{'} z_l} \qquad (2)$$

where z_k and z_l indicate the standardized variables describing the phenomenon under investigation, as observed, respectively, in State k and in State l, and where W is the non-stochastic (NxN) spatial weights matrix. In our case, we use a binary spatial weights matrix. In particular, when two States are neighbors (i.e. they share a common border), the corresponding entry in the matrix is one, and zero otherwise. The elements on the main diagonal are zero by construction, since a State cannot be contiguous to itself[9]. The spatial weights matrix is row standardized, so that neighboring variables are weighted averages of the values in neighboring States (Anselin, 1988). MI allows us to establish a relationship between a phenomenon observed in a given State l, and the same phenomenon observed in contiguous States. Values of MI range from −1 to +1. Negative values indicate negative spatial autocorrelation, and positive values indicate positive spatial autocorrelation. A zero value indicates a random spatial pattern.

The joint analysis of the two indeces yields information on the dynamic process at work at the State level. On the one hand, an increase (decrease) in spatial dependence could be due to the States in each cluster (i.e. each region) becoming more (less) similar in their obesity rates. On the other hand, an increase (decrease) in spatial dependence could also be due to newly formed, extended (reduced)

[9] Our spatial analysis was carried out using the Queen contiguity spatial weights matrix of order 1 (Q1) (see Anselin, 1996). In order to verify the robustness of our results, we use other contiguity matrices. Queen matrices of order 1 and 4 were compared, whereas for k-nearest matrices, the mean of neighbors (4) was used and the standard deviation (2) was added one at a time, and so the orders of the k-nearest matrices were 4, 6, 8 and 10 (Anselin, 1996). The matrices of this type, being contiguity ones, were sequentially selected. Finally, we use as a measure of the distance between a given State and others the inverse of the distance, expressed in km, between the geographical centers of the two States. This distance matrix has an interesting meaning: the increase of distance reduces the strength of ties between a given State and neighboring ones. Note that there is no significant difference between the indices calculated for the different contiguity matrices. In addition, the Spearman rank correlation coefficient calculated for all variables indicated a 90% significant correlation. As the various contiguity matrices were not statistically different, we decided to use the Q1 matrix. We omit these results because they do not add any useful information. Interested readers are welcome to request these results from the authors.

ACCEPTED MANUSCRIPT

clusters emerging during a phase of increased spatial dispersion of obesity rate levels (reduction of Theil Index).

For our data, MI values are all positive and significant at 1%, with the exception of 1990 and 1991 values, with an increase from 0.06 to 0.25 along the sample period. In particular, MI slowly increases, showing that obesity rate levels are related in time and space, and that the spatial diffusion process proceeds slowly over time. In other words, a positive MI indicates that: a high (low) OR observed in a particular State is associated to high (low) OR in contiguous States (positive spatial correlation). In this case, it is likely that unsatisfactory (satisfactory) obesity rate levels tend to spread across States. In addition, Figure 3 shows that MI tends to move in discordance with inequality as measured by TI over time, i.e. a decrease in inequality causes a spatial diffusion effect for obesity rates[10]. This effect leads to the formation of new spatial clusters of States characterized by more intense connection (increase in MI). The correlation between MI and TI is -0.55 over a 22-year sample period. In addition, since dispersion of obesity rate levels moves in the opposite direction of spatial autocorrelation, States with relatively satisfactory (unsatisfactory) obesity rates tend to be located close to others with similar obesity rate levels. Such spatial clusters characterized by satisfactory vs. unsatisfactory obesity rates tend to solidify over time (upward trend of MI).

The default hypothesis that observed levels of obesity rates for each State can be treated as independent does not apply here, so that we can conjecture that some social transmission process is

[10] In fact, if Figure 3 would imply that Theil's index is characterized by a stable trend rather than a downward one, our comments would not be appropriate. To check this, we implement the modified Dickey–Fuller t test (known as the DF-GLS test) proposed by Elliott, Rothenberg, and Stock (1996). Essentially, the test involves fitting a regression of the form: $\Delta y_t = \alpha + \beta y_{t-1} + \delta t + \gamma t^2 + \xi_1 \Delta y_{t-1} + \xi_2 \Delta y_{t-2} + \cdots + \xi_k \Delta y_{t-k} + \varepsilon_t$, where y is the variable analyzed (in our case Theil's index), t and $t^2$are the linear and the quadratic trend respectively, Δy_{t-k} are the time lags of the analyzed variable, and ε_t is the stochastic error term. The number of time lags is determined by the AIC (Akaike information criterion). The DF-GLS test tests the null hypothesis of whether a unit root is present in an autoregressive model. In our case, as shown in the Appendix at the end of the paper, we reject the null hypothesis, and we can conclude that the series is not stationary. Additionally, the test returns the parameters of the estimated equation. In particular, we verify that the linear trend is negative and significant, while the quadratic trend is positive and not significant. This allows us to conclude that the series is characterized by a decreasing linear trend.

at work, that causes a harmonization of obesity rates across nearby territories. The policy implications of these findings are important. In particular, States that are adjacent to others with unsatisfactory obesity rates may be negatively affected even if the current obesity rate is not as bad. In this case, action is needed in terms of countervailing measures such as promotion of healthy lifestyles (e.g., healthy eating, sports activities, etc.), and especially so in border areas with the most critical neighbors, where mobility-driven interaction is most intense. In addition, the cluster of States characterized by satisfactory obesity rates can be taken as a source of good practices, and as a benchmark case for States with worse obesity rate levels.

[Insert Figure 3 about here]

4. Spatial Markov Chains analysis

The transition spatial Markov matrix is calculated, as anticipated, for obesity rates at State level. According to this methodology, the transition of obesity rates takes place between two subsequent time periods. In our analysis, we have twenty-one possible transitions in the period 1990-2011 (e.g., 1990-1991, 1991-1992, …, 2010-2011), and for each couple of years we calculate the number of cases for each class. Its classes (i.e., the cells of the matrix) show the transition probabilities, i.e. the probabilities that a State belonging to class i at time t, ends up belonging to class j in time $t + 1$ (Le Gallo, 2004, Le Gallo and Ertur, 2003). Moreover, as the Spatial Markov Chains analysis deals with the transition from one status to another, it is necessary to categorize the rate of obesity. The use of a continuous variable precludes the use of Spatial Markov Chains analysis which is, to our knowledge, the most analytically compelling way to study the spatial transition dynamics that we analyze in this paper, on the basis of the available dataset. Consequently, it is necessary to proceed in defining classes of obesity rates. As we have 50 US States plus the District of Columbia (n=51),

ACCEPTED MANUSCRIPT

21 years (t-1) from 1990 to 2011, and 5 (K) classes[11], it is possible to obtain, at most, *(51*5*21)=5,355* cases of transitions[12].

Our analysis is preceded by a linear trend control of the obesity rates series of individual US states through the modified Dickey–Fuller *t* test (for more details see footnote 8). The modified Dickey-Fuller Test is a test to verify the stationarity of the obesity rates. Obesity rates are stationary if they do not have trend or seasonal effects. In other words, summary statistics calculated on the obesity rates are consistent over time, like the mean or the variance of the observations. The modified Dickey-Fuller test can be used with serial correlation. This test can handle more complex models than the standard Dickey-Fuller test, and it is also more powerful. The hypotheses for the test are the following: 1) The null hypothesis is that there is no stationarity. 2) The alternate hypothesis differs slightly according to which equation we are using. The basic alternative is that obesity rates are stationary (or trend-stationary). In our case, we could come to wrong results if trends in obesity in the US were not linear throughout the period of analysis. If obesity patterns were characterized by periods of sudden outbursts and others of relative stability, the coefficients of transition might not be stable over time. The modified Dickey–Fuller *t* test rejects the null hypothesis, and we can conclude that the series is not stationary. In particular, we verify that the linear trend is positive and significant for all series', whereas the quadratic trend is not significant (see Table A in Appendix). This allows us to conclude that the series is characterized by a positive linear trend, and not by more complex dynamic patterns; in other words, the quadratic trend term is not statistically significant. Therefore, our analysis is consistent with the structure of the data.

The analysis is conducted using various contiguity matrices, as described in footnote 7. As for the Moran index, also for the Spatial Markov Chains analysis the findings are equivalent, and

[11] The number of classes (= 5) is given by default by the software STARS (Space-Time Analysis of Regional Systems) (Rey and Janikas, 2006), and it is not editable.
[12] With n States, K states and t years, there are $(t\text{-}1)*K*n$ possible cases of transitions.

consequently we only present the results obtained using the Queen contiguity spatial weights matrix of order 1 (Q1)[13].

We report the SMCs results as in Rey (2001). In particular, we define five feasible classes (*K*=5) based upon the observed values of the state variable (obesity rate), with respect to the mean (M) (Schettini et al., 2011). We can accordingly define the following classes:

- Good obesity rate (G), with a characteristic value of the obesity rate lying below the mean up to ¾ of a standard deviation (0.860);
- Fair obesity rate (F), with a characteristic value of the obesity rate lying below the mean between ¾ and ¼ of a standard deviation (0.968);
- Average obesity rate (A), with a characteristic value of the obesity rate equal to 1.058 (the average sample value) plus/minus ¼ of a standard deviation;
- Inadequate obesity rate (I), with a characteristic value of the obesity rate lying above the mean between ¼ and one standard deviation (1.153);
- Bad obesity rate (B), with a characteristic value of the obesity rate lying above the mean between one and 1½ standard deviation (1.406).

Therefore, the five classes can be ordered as follows, from best to worst: G<F<A<I<B. The ascending order reflects the increase in the obesity rates as we move from class to class.

The results of conditioning the transition probabilities on the spatial lag[14] of a given State are reported in Table 2, where column 4 lists the number of cases in each situation. For example, line 8 indicates the transition probability of a State that starts in *t* with an Average level of obesity rate, to

[13] We omit these results because they do not add any significant information. Interested readers are welcome to request them from the authors.

[14] The spatial lag is the average obesity rate of neighboring States. Specifically, the spatial lag is a weighted average, where the weights are represented by the elements of the contiguity matrix.

move to a different obesity rate class in the following year (*t+1*), given that it is surrounded by Fair neighbors. If we consider pairs of consecutive years, there are 80 cases (line 8 column 4) of States in that situation.

Lines 1-5 represent States sitting in neighborhoods with Good obesity rates (G); lines 6-10 represent States sitting in neighborhoods with Fair obesity rates (F); lines 11-15 represent States sitting in neighborhoods with Average obesity rates (A); lines 16-20 represent States sitting in neighborhoods with Inadequate obesity rates (I); finally, lines 21-25 represent States sitting in neighborhoods with Bad obesity rates (B). It is interesting to note that the shaded cells generally deploy the highest values for each line, and as such cells denote the main diagonal, this reveals the presence of inertia: the probability of a State to remain in the same obesity rate class from year to year is relatively high, and in some cases such probability reaches 0.80. For the sake of conciseness, in the remainder of the paper we will use the shorthand expressions "Good (Fair, Average, etc.) States", to denote those States that are categorized as belonging to the G (F, A, etc.) class in a given year, and accordingly for their possible transitions from one class to another, with the respective probabilities as determined by the spatial Markov matrix.

If we focus on States with satisfactory obesity rates (Good or Fair), we observe that the probability of remaining in the favorable condition is:

- high for Fair States sitting in neighborhoods with Good obesity rates. In particular, it is equal to 0.895 (sum of the cells up to F, line 2). The same probability is equal to 0.929 for the Good States (sum of the cells up to F, line 1). Moreover, Fair States have a probability of 0.184 of improving if they are surrounded by Good States (the cell in correspondence of G, line 2);
- high for both Good and Fair States sitting in neighborhoods with Fair obesity rates: 0.979 for Good States (sum of the cells up to F, line 6) and 0.719 for Fair States (sum of the cells

ACCEPTED MANUSCRIPT

up to F, line 7). In particular, Good States have a probability of 0.274 of worsening their status if they are surrounded by Fair States (the cell in correspondence of F, line 6);

— high for both Good and Fair States sitting in neighborhoods with Average obesity rates (A): 0.844 for Good States (sum of the cells up to F, line 11) and 0.684 for Fair States (sum of the cells up to F, line 12). Good States have a probability of 0.125 (0.156) of worsening their status if they are surrounded by Fair (Average) States (the cell in correspondence of F (A), line 11). Also, Fair States have a probability of 0.217 of worsening their status if they are surrounded by Average States (the cell in correspondence of A, line 12);

— finally, it is interesting to note that, in the case of neighborhoods characterized by unsatisfactory obesity rates (Inadequate or Bad), both Good and Fair States have a high probability of worsening their status. In particular, in the case of neighborhoods with Inadequate obesity rates (I), Good States have a probability of 0.5 (0.1) of worsening their status (the cell in correspondence of F (B), line 16). Moreover, Fair States have a probability of 0.58 of worsening their status (the cell in correspondence of A, line 17). In the case of neighborhoods with Bad obesity rates (B), Good (Fair) States have a probability of 0.2 (1) of worsening their status (the cell in correspondence of F (A), line 21 (22)).

We now consider States starting off with Inadequate or Bad obesity rates (in the year *t*), that is, States with above-average obesity rates (Inadequate or Bad), but sitting in neighborhoods with Good or Fair obesity rates (lines 4, 5, 9, 10). In this case, we can determine the probability of obesity rates to remain above average or to fall below average in the subsequent period. In particular, we note that Good States have a positive effect on Inadequate and Bad States: the probability that the States with Inadequate obesity rates pass into an Average obesity rate class is equal to 1 (the cell in correspondence of A, line 4). Also, Bad States have a probability of 0.25 of improving their status (the cells in correspondence of I and F, line 5). Moreover, Fair States have a positive effect on Inadequate and Bad States: the probability that States with Inadequate obesity

ACCEPTED MANUSCRIPT

rates pass into an Average (Fair) obesity rate class is equal to 0.314 (0.143) (the cell in correspondence of A (F), line 9). And Bad States have a probability of 0.33 (0.44) of improving their status (the cells in correspondence of I (F), line 10).

When States starting with Inadequate or Bad obesity rates are surrounded by States with similar obesity rates (Inadequate and Bad States, respectively) (lines 19, 20, 24 and 25), we observe that:

- Inadequate States, if surrounded by other Inadequate States, have a higher probability to worsen their obesity rate (0.207, in correspondence of B) than to improve it (0.107, in correspondence of A) (line 19). Bad States surrounded by Inadequate States have a probability of 0.25 to improve into Inadequate States, and of 0.03 to improve into Average States, respectively (in correspondence of I and A, line 20).

- Inadequate States, if surrounded by Bad States, have a probability of 0.333 (0.111) to worsen (improve) their obesity rate (in correspondence of B (A), line 24). On the contrary, Bad States surrounded by other Bad States have a 0.242 probability to improve their obesity rate (in correspondence of I, line 25).

Finally, if Inadequate and Bad States are surrounded by Average States, we observe that: Inadequate States have a probability of about 0.129 (0.271 and 0.035) to worsen (improve) their obesity rate (in correspondence of B (A and F, respectively), line 14), whereas Bad States have a probability of 0.029 to improve their obesity rate (in correspondence of A, line 15).

To sum up, we observe that States with satisfactory obesity rates (Good or Fair) are closely linked in terms of proximity effects: Good States affect Fair States and vice versa, and this influence is positive because it reciprocally improves obesity rates. In addition, these States (Good and Fair), if surrounded by Average States, worsen their obesity rates to a limited amount. Moreover, Good and Fair States are negatively affected by States with Inadequate or Bad obesity rates. Finally, Inadequate and Bad States are positively affected by Good and Fair States.

ACCEPTED MANUSCRIPT

We now also consider the ergodic distribution[15] that can be interpreted as the long-run distribution of obesity rates at State level. Additional insights about the relationship between a State's transition probabilities and the obesity rate class of its spatial lag can be gained by considering the ergodic distributions implied by each of the conditional transition matrices from Table 2. Five different ergodic state vectors are reported in Table 3.

Like the initial distributions, the long-run distributions are biased. Indeed, when States are surrounded by neighbors with above-average obesity rates (Inadequate or Bad), the final distribution is more and more skewed upwards: the probability to maintain an unsatisfactory obesity rate in the long run is high (Table 3, columns I and B). Alternatively, when States are surrounded by neighbors with under-average obesity rates (Good or Fair), the ergodic distribution is more and more negatively skewed: the probability to maintain a satisfactory obesity rate in the long run is very high (Table 3, columns G and F).

[Insert Table 2 about here]

[Insert Table 3 about here]

In Tables 4 and 5, we report the information extracted from the results presented in Table 2. In particular, Table 4 shows the probability of a State to stay in the same class of obesity rates, independently of its neighborhood (Schettini et al., 2011). In this case, we observe that such probability is high for Good, Average, and Bad classes, and respectively equal to 0.6106, 0.5016 and 0.5806; less high for the Inadequate class (0.4472), while the lowest probability is registered for the Fair class (0.4064).

[Insert Table 4 about here]

[15] "The ergodic distribution should be viewed as a "thought experiment" that illustrates how space may influence transition dynamics, rather than as a guide to what would transpire in reality" (Rey, 2011). The ergodic distribution delivered by the software is computed for each of the five transition matrices. For more details on the ergodic distribution concept, see Rey (2001) and Le Gallo (2004).

Following Schettini et al. (2011), we count in Table 5 all the cases of States whose neighbors sit in better obesity rate classes (Table 5**,** first row, column X) and, among these, we count the cases of States that improve their class (Table5**,** first row, column Y). At this point, we calculate the probability of moving to better obesity rate classes, given that the State is surrounded by neighbors with better obesity rate levels (first row, column Y/X in Table 5). The same method is applied to the cases of worsening obesity rate class in the second row of the matrix.

The calculations reported in Table 5 show that: 1) if a State is surrounded by neighbors with better (worse) obesity rates, it has a probability of about 0.5754 (0.6526) to improve (worsen) its obesity rate; 2) the probability of improving the obesity rate is lower than that of worsening it. We can thus conclude that the *pull effect* (i.e., the positive impact of neighbors with satisfactory obesity rates upon the improvement of a State's obesity rate) is lower than the *drag effect* (i.e., the negative impact of neighbors with unsatisfactory obesity rates upon the worsening of a State's obesity rate) (see Schettini et al., 2011).

[Insert Table 5 about here]

5. Discussion

In this paper, we have found that, in the case of US data at State level for the period 1990-2011, the dynamics of obesity rates are subject to quite significant proximity effects, and are therefore compatible with the hypothesis of social transmission processes that influence obesity-related attitudes and choices to an extent that yields a traceable aggregate impact. In lack of an explicit micro modeling of such processes and of its empirical validation, one cannot currently go beyond the assessment of the compatibility of the evidence with the stated hypothesis, but it is difficult to think of an alternative causal mechanism, also in view of the strength of the reported effects at work. This is clearly a strong motive to start looking more closely into this heavily under-researched, yet very policy-relevant area.

Our analysis therefore illustrates how the idea of an 'obesity epidemics' is not just a powerful metaphor, but may be a worrisome reality, and that the 'infectious' agent is in this case not a viral one, but most likely a complex bundle of transmissible social cues, part of which already singled out in the literature cited in the introduction, that has studied in detail the insurgence of obesity in specific social environments such as adolescent groups. Developing a more precise and detailed understanding of such cues, and possibly even a taxonomy of their complementary or antagonistic functioning, proves to be vital to design effective countervailing actions and policies. Research is still at an early stage, and far from systematic. However, the fact that the spatial element turns out to be crucial for social transmission should be considered more carefully, and more effectively accounted for in the design of such policies.

The analysis of the US case shows a nuanced, complex picture. For instance, States characterized by low obesity rates have good chances to preserve their satisfactory condition. However, for States with fair but not optimally low levels of obesity rates, being surrounded by States with considerably higher obesity rates could be disruptive to some extent. As States with high obesity rates tend to cluster together at a spatial level, the risk of a perverse lock-in situation becomes substantial, and a real improvement can only be obtained through a joint effort at inter-State level rather than through an isolated initiative by a single State. This could be achieved, in particular, through coordinated inter-State design of anti-obesity policy measures for particularly sensitive target groups such as children, youths, and socio-economically deprived categories. These are merely a couple examples of the insights that derive from our analysis.

There are many possible directions in which this preliminary analysis can be extended. In the first place, it would be of interest to extend the analysis on US data to an analysis of other countries

ACCEPTED MANUSCRIPT

characterized by similar levels of socio-economic development, by means of data at a comparable level of spatial aggregation. In case the size and structure of the proximity effects would be different for other countries, this could give us some hints on the nature of the social transmission effects at work in either case.

In this paper, we have tested a proximity effects model in its simplest form, disregarding the action of other intervening variables that could influence social transmission across States, as a necessary first step to assess the relevance of the proximity factor. This is at the same time the strength and the main limitation of our study. Clearly, there could be several variables of interest to consider in a generalized model where proximity interacts with the geographical distribution of factors such as income, educational level or political orientation, to name a few obvious ones. Likewise, we considered a simple proximity criterion in terms of geographical contiguity, without taking into account the strength of interaction between neighboring States (as measured, for instance, by the size of inter-State commuter flows). These are examples of further promising directions along which the analysis presented in our paper could be extended in future research.

Another interesting direction of development would be a micro-foundation of our analysis on the basis of a specific, explicit micro-model of the social transmission dynamics that reproduces the aggregate dynamics found in our analysis. Consequently, it would be very interesting to compare the aggregate dynamics generated by alternative micro-mechanisms of social transmission, to select the one which provides a better replication.

Finally, the possibility of working with more fine-grained data (for instance at county level, even if for a sub-national or regional universe) would probably provide an even better insight on how the proximity effects actually function, the main issue being the availability of long enough time series at county/province level. We hope that all these promising lines of research will be pursued in the near future, in the interest of a more effective, socially beneficial tackling of the obesity epidemics.

ACCEPTED MANUSCRIPT

Appendix

[Insert Table A about here]

References

Agovino M, Crociata A, Sacco PL (2016) Location matters for pro-environmental behavior: A spatial Markov Chains approach to proximity effects in differentiated waste collection. Annals of Regional Science 56, 295-315.

Anselin L (1988) Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geographical Analysis 20, 1-17.

Anselin L (1996) Interactive techniques and exploratory spatial data analysis. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW (Eds.), Geographic Information Systems: Principles, Techniques, Management and Applications. John Wiley, New York, pp. 251–56.

Anselin L (2010) Thirty years of spatial econometrics. Papers in Regional Science 89, 3-25.

Anselin L, Bera A (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles DEA (eds) Handbook of applied economic statistics. Marcel Dekker, New York

Arbia G (2005) Spatial econometrics, statistical foundations and applications to regional convergence. Springer, Berlin

Arbia G, Basile R, Mirella S (2002) Regional convergence in Italy 1951–199: a spatial econometric perspective. ISAE working papers 29, ISAE—Institute for Studies and Economic Analyses (Rome, ITALY).

ACCEPTED MANUSCRIPT

Arcaya M, Brewster M, Zigler CM, Subramanian SV (2012) Area variations in health: a spatial multilevel modeling approach. Health and Place, 18(4), 1-24.

Arcaya MC, Tucker-Seeley RD, Kim R, Schnake-Mahl A, So M, Subramanian SV (2016) Research on neighborhood effects on health in the United States: A systematic review of study characteristics. Social Science and Medicine 168, 16-29.

Baum CL II, Ruhm CJ (2009) Age, socioeconomic status and obesity growth. Journal of Health Economics 28, 635-48.

Chou SY, Grossman M, Saffer H (2004) An economic analysis of adult obesity: Results from the Behavioral Risk factor Surveillance System. Journal of Health Economics 23, 565-87.

Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. New England Journal of Medicine 357, 370-9.

Cohen DA, Finch BK, Bower A, Sastry N (2006) Collective efficacy and obesity: The potential influence of social factors on health. Social Science and Medicine 62, 769-78.

Cohen-Cole E, Fletcher JM (2008) Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. Journal of Health Economics 27, 1382-87.

Contaldo F, Pasanisi F (2004) Obesity epidemics: Secular trend or globalization consequence? Beyond the interaction between genetic and environmental factors. Clinical Nutrition 23, 289-91.

Contaldo F, Pasanisi F (2005) Obesity epidemics: Simple or simplistic answers? Clinical Nutrition 24, 1-4.

Cooksey-Stowers K, Schwartz MB, Brownell KD (2017) Food swamps predict obesity rates better than food deserts in the United States. International Journal of Environmental Research and Public Health 14, 1366, doi:10.3390/ijerph14111366.

ACCEPTED MANUSCRIPT

Dammann KW, Smith C (2009) Factors affecting low-income women's food choices and the perceived impact of dietary intake and socioeconomic status on their health and weight. Journal of Nutrition Education and Behavior 41, 242-53.

Davis B, Carpenter C (2009) Proximity of fast-food restaurants to schools and adolescent obesity. American Journal of Public Health 99, 505-10.

de la Haye K. Robins G, Mohr P, Wilson C (2010) Obesity-related behaviors in adolescent friendship networks. Social Networks 32, 161-7.

de la Haye K. Robins G, Mohr P, Wilson C (2011) How physical activity shapes, and is shaped by, adolescent friendships. Social Science and Medicine 73, 719-28.

Dodson EA, Fleming C, Boehmer TK, Haire-Joshu D, Luke DA, Brownson RC (2009) Preventing childhood obesity through State policy: Qualitative assessment of enablers and barriers. Journal of Public Health Policy 30, S161-76.

Dragone D, Savorelli L (2012) Thinness and obesity: A model of food consumption, health concerns, and social pressure. Journal of Health Economics 31, 243-56.

Ejima K, Aihara K, Nishiura H (2013) Modeling the obesity epidemic: Social contagion and its implications for control. Theoretical Biology and Medical Modelling 10:17, doi 10.1186/1742-4682-10-17.

Elfhag K, Morey LC (2008) Personality traits and eating behavior in the obese: poor self-control in emotional and external eating but personality assets in restrained eating. Eating Behaviors 9, 285-93.

Elliott GR, Rothenberg TJ, Stock JH (1996) Efficient tests for an autoregressive unit root. Econometrica 64, 813–36.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Fitzgerald A, Fitzgerald N, Aherne C (2012) Do peers matter? A review of peer and/or friends' influence on physical activity among American adolescents. Journal of Adolescence 35, 941-58.

Gezici F, Hewings GJ D (2007) Spatial Analysis of Regional Inequalities in Turkey, European Planning Studies, 15:3, 383-403.

Ghosh-Dastidar B, Cohen D, Hunter G, Zenk SN, Huang C, Beckman R, Dubowitz T (2014) Distance to store, food prices, and obesity in urban food deserts. American Journal of Preventive Medicine 47, 587-95.

Griffith DA, Paelinck JHP (2007) An equation by any other name is still the same: On spatial econometrics and spatial statistics. Annals of Regional Science 41, 209-27

Halliday TJ, Kwak S (2009) Weight gain in adolescent and their peers. Economics and Human Biology 7, 181-90.

Hawkes C, Smith TG, Jewell J, Wardle J, Hammond RA, Friel S, Thow AM, Kain J (2015) Smart food policies for obesity prevention. The Lancet 385, 2410-21.

Heymsfield SB, Wadden TA (2017) Mechanisms, pathophysiology, and management of obesity. The New England Journal of Medicine 376, 254-266.

Hruschka DJ (2012) Do economic constraints on food choice make people fat? A critical reviewer of two hypotheses for the poverty-obesity paradox. American Journal of Human Biology 24, 277-85.

Inglis V, Ball K, Crawford D (2009) Does modifying the household food budget predict changes in the healthfulness of purchasing choices among low- and high-income women? Appetite 52, 273-79.

Jetter KM, Cassady DL (2006) The availability and cost of healthier food alternatives. American Journal of Preventive Medicine 30, 38-44.

ACCEPTED MANUSCRIPT

Lawrence NS, Hinton EC, Parkinson JA, Lawrence AD (2012) Nucleus accumbens response to food cues predicts subsequent snack consumption in women and increased body mass index in those with reduced self-control. NeuroImage 63, 415-22.

Le Gallo J (2004) Space-time analysis of GDP disparities among European regions: A Markov chains approach. International Regional Science Review 27, 138-63.

Le Gallo, J., Ertur, C., 2003. Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980–1995. Pap. Reg. Sci. 82 (2), 175–201.

Lin BH, Ver Ploeg M, Kasteridis P, Yen ST (2014) The roles of food prices and food access in determining food purchases of low-income households. Journal of Policy Modeling 36, 938-52.

Ludwig J, Sanbonmatsu L, Gennetian L, Adam E, Duncan GJ, Katz LF, Kessler RC, Kling JR, Tessler Lindau S, Whitaker RC, McDade TW (2011) Neighborhoods, obesity anddiabetes – A randomized social experiment. The New England Journal of Medicine 365, 1509-19.

Madan A, Pentland A (2009) Modeling social diffusion phenomena using reality mining. AAAI Spring Symposium: Human behavior modeling, 43-48, http://www.aaai.org/Papers/Symposia/Spring/2009/SS-09-04/SS09-04-008.pdf.

Madan A, Motoru ST, Lazer D, Pentland A (2010) Social sensing: Obesity, unhealthy eating and exercise in face-to-face networks. Wireless Health 2010, Association for Computer Machinery, San Diego CA http://dspace.mit.edu/openaccess-disseminate/1721.1/65376.

Minet Kinge J, Morris S (2010) Socioeconomic variation in the impact of obesity on health-related quality of life. Social Science and Medicine 71, 1864-71.

Morland KB, Evenson KR (2009) Obesity prevalence and the local food environment. Health and Place 15, 491-5.

Mullan B, Ntoumanis N, Thøgersen-Ntoumani C, Lipp OV (2017) "It's a bit more complicated than that". A broader perspective on determinants of obesity. Behavioral and Brain Sciences, forthcoming, e124, doi 10.1017/S0140525X16001497.

NCD-RisC [NCD Risk Factor Collaboration] (2016) Trends in adult body-ass index in 200 countries from 1975 to 2014: A pooled analysis of 1698 population-based measurement studies with 19.2 million participants. The Lancet 387, 1377-96.

Nelson GD, Rae A (2016) An economic geography of the United States: From commutes to megaregions. PLoS ONE 11(11): e0166083, doi 10.1371/journal.pone.0166083.

Nettle D, Andrews C, Bateson M (2017) Food insecurity as a driver of obesity in humans: The insurance hypothesis. Behavioral and Brain Sciences, forthcoming, e105, doi S0140525X16000947.

Renna F, Grafova IB, Thakur N (2008) The effect of friends on adolescent body weight. Economics and Human Biology 6, 377-87.

Rey SJ (2001) Spatial empirics for economic growth and convergence. Geographical Analysis 33, 195-214.

Rey SJ, Janikas MV (2006) STARS: Space-time analysis of regional systems. Geographical Analysis 38, 67-86.

Rohilla Shalizi C, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. Sociological Methods and Research 40, 211-39.

Ruhm CJ (2012) Understanding overeating and obesity. Journal of Health Economics 31, 781-96.

Sacco PL (2017) Obesity as self-regulation failure: A 'disease of affluence' that selectively hits the less affluent? Behavioral and Brain Sciences, forthcoming, e127, doi S0140525X16001527.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Santonja FJ, Villanueva RJ, Jódar L, Gonzalez-Parra G (2010) Mathematical modelling of social obesity epidemic in the region of Valencia, Spain. Mathematical and Computer Modelling of Dynamical Systems 16, 23-34.

Scharoun-Lee M, Adair LS, Kaufman JS, Gordon-Larsen P (2009) Obesity, race/ethnicity and the multiple dimensions of socioeconomic status during the transition to adulthood: A factor analysis approach. Social Science and Medicine 68, 708-16.

Schettini D, Azzoni CR, Paez A (2011) Neighborhood and efficiency in manufacturing in Brazilian regions: A spatial Markov chain analysis. International Regional Science Review 34, 397-418.

Shahar D, Shai I, Vardi H, Shahar A, Fraser D (2005) Diet and eating habits in high and low socioeconomic groups. Nutrition 21, 559-66.

Shoham DA, Hammond R, Rahmandad H, Wang Y, Hovmand P (2015) Modeling social norms and social influence in obesity. Current Epidemiology Reports 2, 71-9.

Sniehotta FF, Scholz U, Schwarzer R (2005) Bridging the intention-behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychology and Health 20, 143-60.

Sturm R (2007) Increases in morbid obesity in the USA: 2000-2005. Public Health 121, 492-6.

Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, Gortmaker SL (2011) The global obesity pandemic: Shaped by local drivers and global environments. The Lancet 378, 804-14.

Trogdon JG, Nonnemaker J, Pais J (2008) Peer effects in adolescent overweight. Journal of Health Economics 27, 1388-99.

Valente TW, Fujimoto K, Chou CP, Spruijt-Metz D (2009) Adolescent affiliation and adiposity: A social network analysis of friendship and obesity. Journal of Adolescent Health 45, 202-4.

ACCEPTED MANUSCRIPT

Walker RE, Keane CR, Burke JG (2010) Disparities and access to healthy food in the United States: A review of the food deserts literature. Health and Place 16, 876-84.

Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M (2011) Health and economic burdern of the projected obesity trends in the USA and the UK. The Lancet 378, 815-25.

Wolch J, Jerrett M, Reynolds K, McConnell R, Chang R, Dahmann N, Brady K, Gilliland F, Su JG, Berhane K (2011) Childhood obesity and proximity to urban parks and recreational resources: A longitudinal cohort study. Health and Place 17, 207-14.

Yue W, Zhang Y, Ye X, Cheng Y, Leipnik MR (2014) Dynamics of multi-scale intra-provincial regional inequality in Zhejiang, China. Sustainability 6(9), 5763-84, doi 10.3390/su6095763.

Zhang J, Tong L, Lamberson PJ, Durazo-Arvizu RA, Luke A, Shoham DA (2015) Leveraging social influence to address overweight and obesity using agent-based models: The role of adolescent social networks. Social Science and Medicine 125, 203-13.

Zhang Q, Wang Y (2004) Socioeconomic inequality of obesity in the United States: Do gender, age and ethnicity matter? Social Science and Medicine 58, 1171-80.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1. A Spatial Markov matrix

		status at time (t+1)				
status at time t	***status of the neighbors***	***G***	***F***	***A***	***I***	***B***
G	***G***	$p_{GG\|G}$	$p_{GF\|G}$	$p_{GA\|G}$	$p_{GI\|G}$	$p_{GB\|G}$
F		$p_{FG\|G}$	$p_{FF\|G}$	$p_{FA\|G}$	$p_{FI\|G}$	$p_{FB\|G}$
A		$p_{AG\|G}$	$p_{AF\|G}$	$p_{AA\|G}$	$p_{AI\|G}$	$p_{AB\|G}$
I		$p_{IG\|G}$	$p_{IF\|G}$	$p_{IA\|G}$	$p_{II\|G}$	$p_{IB\|G}$
B		$p_{BG\|G}$	$p_{BF\|G}$	$p_{BA\|G}$	$p_{BI\|G}$	$p_{BB\|G}$
G	***F***	$p_{GG\|F}$	$p_{GF\|F}$	$p_{GA\|F}$	$p_{GI\|F}$	$p_{GB\|F}$
F		$p_{FG\|F}$	$p_{FF\|F}$	$p_{FA\|F}$	$p_{FI\|F}$	$p_{FB\|F}$
A		$p_{AG\|F}$	$p_{AF\|F}$	$p_{AA\|F}$	$p_{AI\|F}$	$p_{AB\|F}$
I		$p_{IG\|F}$	$p_{IF\|F}$	$p_{IA\|F}$	$p_{II\|F}$	$p_{IB\|F}$
B		$p_{BG\|F}$	$p_{BF\|F}$	$p_{BA\|F}$	$p_{BI\|F}$	$p_{BB\|F}$
G	***A***	$p_{GG\|A}$	$p_{GF\|A}$	$p_{GA\|A}$	$p_{GI\|A}$	$p_{GB\|A}$
F		$p_{FG\|A}$	$p_{FF\|A}$	$p_{FA\|A}$	$p_{FI\|A}$	$p_{FB\|A}$
A		$p_{AG\|A}$	$p_{AF\|A}$	$p_{AA\|A}$	$p_{AI\|A}$	$p_{AB\|A}$
I		$p_{IG\|A}$	$p_{IF\|A}$	$p_{IA\|A}$	$p_{II\|A}$	$p_{IB\|A}$
B		$p_{BG\|A}$	$p_{BF\|A}$	$p_{BA\|A}$	$p_{BI\|A}$	$p_{BB\|A}$
G	***I***	$p_{GG\|I}$	$p_{GF\|I}$	$p_{GA\|I}$	$p_{GI\|I}$	$p_{GB\|I}$
F		$p_{FG\|I}$	$p_{FF\|I}$	$p_{FA\|I}$	$p_{FI\|I}$	$p_{FB\|I}$
A		$p_{AG\|I}$	$p_{AF\|I}$	$p_{AA\|I}$	$p_{AI\|I}$	$p_{AB\|I}$
I		$p_{IG\|I}$	$p_{IF\|I}$	$p_{IA\|I}$	$p_{II\|I}$	$p_{IB\|I}$
B		$p_{BG\|I}$	$p_{BF\|I}$	$p_{BA\|I}$	$p_{BI\|I}$	$p_{BB\|I}$
G	***B***	$p_{GG\|B}$	$p_{GF\|B}$	$p_{GA\|B}$	$p_{GI\|B}$	$p_{GB\|B}$
F		$p_{FG\|B}$	$p_{FF\|B}$	$p_{FA\|B}$	$p_{FI\|B}$	$p_{FB\|B}$
A		$p_{AG\|B}$	$p_{AF\|B}$	$p_{AA\|B}$	$p_{AI\|B}$	$p_{AB\|B}$
I		$p_{IG\|B}$	$p_{IF\|B}$	$p_{IA\|B}$	$p_{II\|B}$	$p_{IB\|B}$
B		$p_{BG\|B}$	$p_{BF\|B}$	$p_{BA\|B}$	$p_{BI\|B}$	$p_{BB\|B}$

Notes: G is a good health status; F is a fair health status; A is an average health status; I is an inadequate health status; B is a bad health status. Finally, p is the transition probability.

Table 2. SMCs matrix

	t					***(t+1)***		
Line	neighborhood condition		num. cases	**G**	**F**	**A**	**I**	**B**
1	G	G	14	0.429	**0.500**	0.071	0	0

ACCEPTED MANUSCRIPT

2	F		38	0.184	**0.711**	0.079	0.026	0
3	A		12	0	0.333	**0.583**	0.083	0
4	I		1	0	0	**1.000**	0	0
5	B		4	0	0.250	0	0.250	**0.500**
6	G	F	95	**0.705**	0.274	0.011	0.011	0
7	F		135	0.200	**0.519**	0.267	0.015	0
8	A		80	0.037	**0.450**	0.362	0.150	0
9	I		35	0.029	0.143	0.314	**0.429**	0.086
10	B		9	0	0	**0.444**	0.333	0.222
11	G	A	32	**0.719**	0.125	0.156	0	0
12	F		60	0.117	**0.567**	0.217	0.083	0.017
13	A		107	0	0.206	**0.551**	0.224	0.019
14	I		85	0	0.035	0.271	**0.565**	0.129
15	B		34	0	0	0.029	0.265	**0.706**
16	G	I	10	0.400	**0.500**	0	0	0.100
17	F		17	0.118	0.235	**0.588**	0.059	0
18	A		41	0.024	0.122	**0.512**	0.220	0.122
19	I		121	0	0	0.107	**0.686**	0.207
20	B		99	0	0	0.030	0.253	**0.717**
21	G	B	5	**0.800**	0.200	0	0	0
22	F		1	0	0	**1.000**	0	0
23	A		6	0	0.167	**0.500**	0	0.333
24	I		28	0	0	0.111	**0.556**	0.333

ACCEPTED MANUSCRIPT

25	B	32	0	0	0	0.242	**0.758**

Note: the largest value in each row is presented in boldface. Shaded cells indicate permanence in the same class across years.

Table 3. Ergodic health status (obesity rates) distributions

Lag	G	F	A	I	B
G	0.218	0.566	0.183	0.033	0
F	0,331	0,645	0,024	0	0
A	0.078	0.187	0.298	0.283	0.154
I	0.013	0.033	0.156	0.419	0.378
B	0	0.017	0.103	0.310	0.569

Table 4. Probability of staying in the same health status class

probability	G	F	A	I	B
	0.6106	0.4064	0.5016	0.4472	0.5806

Table 5. Summary of SMCs analysis

	Cases of States with better neighbors	Cases of States with better neighbors and that got better	Probability of getting better with better neighbors
Getting	***X***	***Y***	***Y/X***

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

better			
	179	**103**	**0.5754**
	Cases of States with worse neighbors	Cases of States with worse neighbors and that got worse	Probability of getting worse with worse neighbors
Getting worse	***X***	***Y***	***Y/X***
	190	**124**	**0.6526**

Table A. Dickey–Fuller t test by State

macro-area	Federal States	linear trend (t)	quadratic trend (t^2)	AR order[+]	t-test[++]
		coefficients			
Midwestern United States	Illinois	1.67841***	−0.0110393	1	-0.672
Midwestern United States	Indiana	0.919493***	−0.00407325	1	-1.56394
Midwestern United States	Iowa	1.18482***	−0.00730314	1	-0.27354

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Midwestern United States	Kansas	0.409764***	0.00658264	1	-1.78021
Midwestern United States	Michigan	0.697166***	0.000937602	1	-1.53066
Midwestern United States	Minnesota	0.473519*	−0.00645977	1	-2.09466
Midwestern United States	Missouri	0.680966***	−0.00381120	1	-0.43215
Midwestern United States	Nebraska	1.13167***	−0.00876694	1	-1.65956
Midwestern United States	North Dakota	1.16795***	−0.00602061	1	-1.2365
Midwestern United States	Ohio	1.30019***	−0.000150514	1	-1.00231
Midwestern United States	South Dakota	0.833933***	0.0150225	2	-1.09058
Midwestern United States	Wisconsin	0.80166***	0.000197465	2	-1.45466
Northeastern United States	Connecticut	1.2703***	−0.0250174	3	0.10692
Northeastern United States	Maine	0.757472***	0.00661722	1	-2.63349
Northeastern United States	Massachusetts	0.8171***	0.00561197	1	-1.15637
Northeastern United States	New Hampshire	0.598118***	0.0113037	1	-1.98293
Northeastern United States	New Jersey	0.953746***	−0.0104614	1	-2.48693
Northeastern United States	New York	0.600833***	−0.00179717	1	-2.68689
Northeastern United States	Pennsylvania	0.937093***	0.00155553	1	-2.06039
Northeastern United States	Rhode Island	0.494385***	0.0106762	2	-1.00453
Northeastern United States	Vermont	0.856386***	−0.00270456	2	-0.14741
Southern United States	Alabama	1.32308***	−0.0150436	2	-0.40362
Southern United States	Arkansas	2.37016***	−0.00263765	1	-2.91242
Southern United States	Delaware	0.338284***	0.0483232	2	-0.79607
Southern United States	District of Columbia	0.643378***	−0.0125798	1	-1.23401
Southern United States	Florida	0.939014*	0.0109273	1	2.3412
Southern United States	Georgia	1.02445***	−0.0079124	1	-2.26507
Southern United States	Kentucky	1.30142***	−0.00699187	1	-0.34456
Southern United States	Louisiana	1.32664***	0.00142481	1	-2.13671
Southern United States	Maryland	0.856098***	−0.00344264	1	-1.21905
Southern United States	Mississippi	0.85334***	0.00426413	1	-0.35395
Southern United States	North Carolina	0.833153***	−0.00480519	2	-0.34567
Southern United States	Oklahoma	0.849306***	0.0141826	1	-0.34567
Southern United States	South Carolina	0.854728***	0.00958182	2	-2.90506
Southern United States	Tennessee	1.2683***	0.00648778	1	1.14466

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

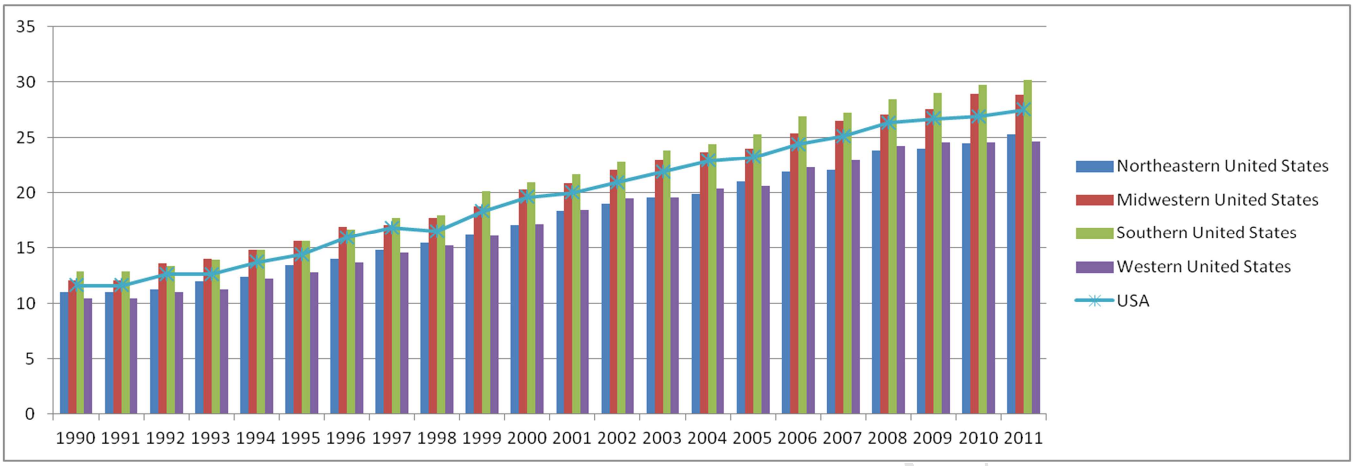
Southern United States	Texas	0.809638***	−0.00719547	1	-1.2249
Southern United States	Virginia	1.4968***	−0.0214447	2	1.25369
Southern United States	West Virginia	0.925102***	−0.00728584	1	-1.4462
Western United States	Alaska	1.2422***	−0.0255245	2	-1.75123
Western United States	Arizona	0.616149	0.0279536	1	0.051538
Western United States	California	1.57395***	−0.0240875	2	0.404792
Western United States	Colorado	1.17439***	−0.00813276	2	-2.49871
Western United States	Hawaii	0.399789***	0.00317707	1	-2.212
Western United States	Idaho	1.07581***	−0.00126886	1	-1.34678
Western United States	Montana	0.732644***	−0.00227221	1	-2.94811
Western United States	Nevada	0.480952***	−0.00427519	1	-1.05287
Western United States	New Mexico	0.819933***	−0.00727706	1	-2.7627
Western United States	Oregon	1.07318***	−0.0160441	1	-2.83707
Western United States	Utah	0.759531***	−0.00923329	1	-1.57117
Western United States	Washington	1.3748***	−0.00793776	1	-2.00065
Western United States	Wyoming	0.220328*	0.00654693	2	-0.2317

*Note: [+]The number of time lags is determined by the AIC (Akaike information criterion); [++]t-test never rejects the null hypothesis of unit root; *** - significant at 1%, ** - significant at 5%, * - significant at 10%..*

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Figure 1. Trend of obesity rate by regional divisions, 1990-2011



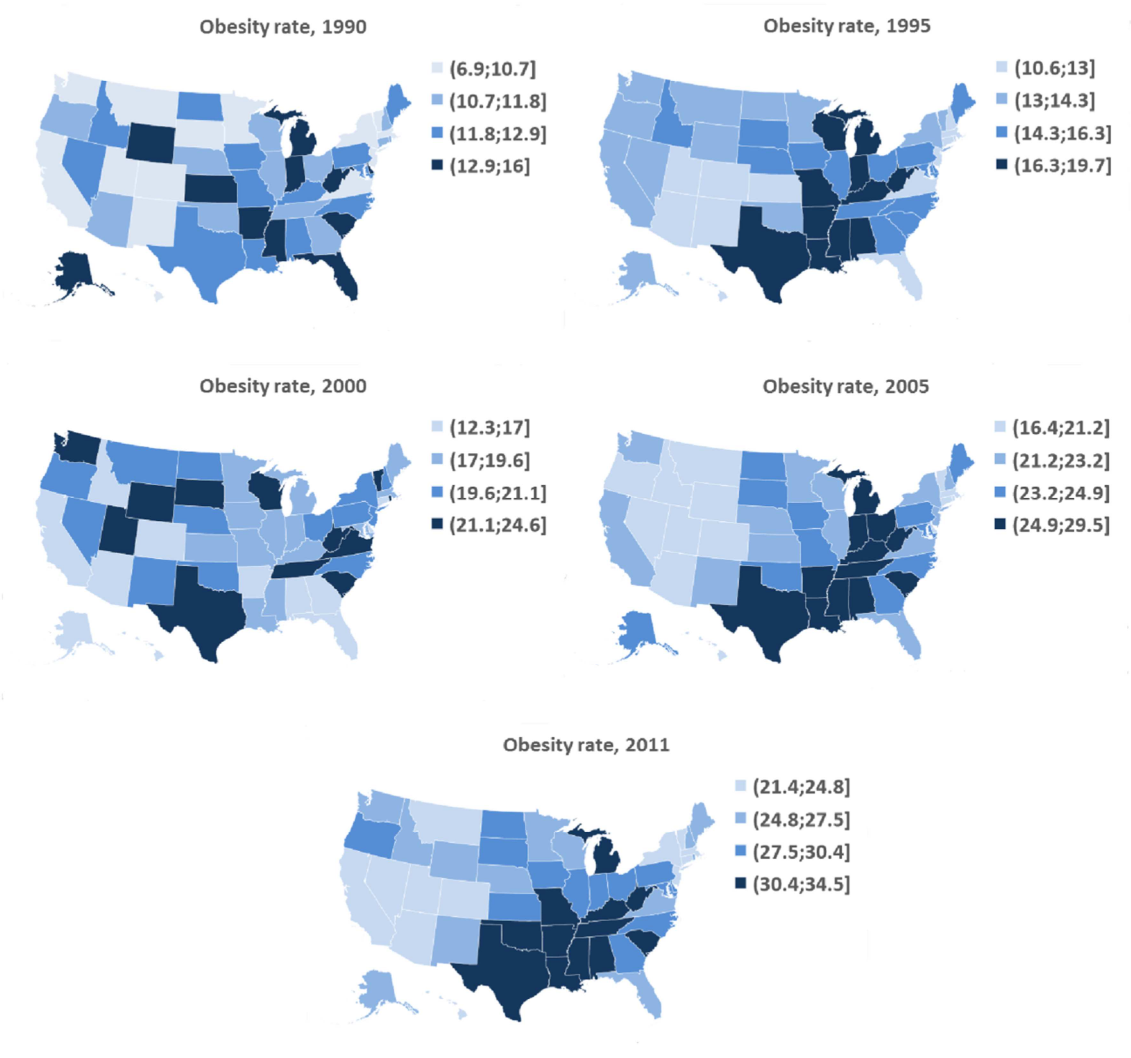
Source: our elaboration on America's Health Rankings data.

Note: Northeastern United States (namely, Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont) ; Midwestern United States (namely, Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin); Southern United States (namely, Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina , Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia); Western United States(namely, Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

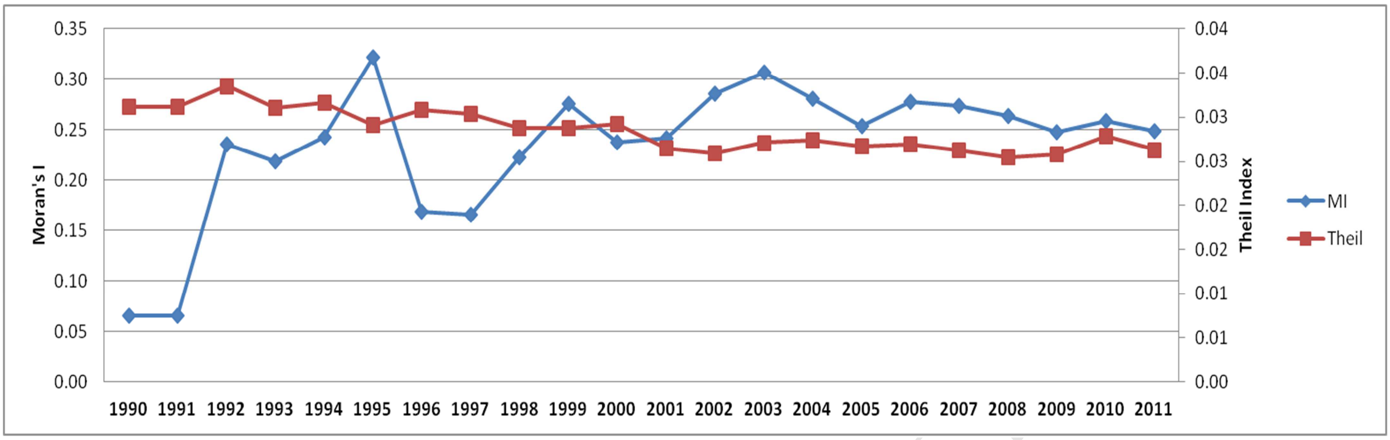
Figure 2. Quartile maps of obesity rates, 1990, 1995, 2000, 2011



Source: our elaboration on America's Health Rankings data.

ACCEPTED MANUSCRIPT

Figure 3. Total inequality and spatial autocorrelation, 1990-2011



Source: our elaboration on America's Health Rankings data.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

- We study the presence of spatial proximity effects in US State-level obesity data
- We propose a spatial approach based on combining Moran and Theil indexes
- We find that the evidence is consistent with an assumption of social contagion
- It may be appropriate to speak of a socially transmitted 'obesity epidemics'

ACCEPTED MANUSCRIPT